Scuola Normale Superiore

Ph.D. Thesis in Chemistry

Understanding the working mechanism of natural and engineered proteins for cell regulation and biotechnological applications: a molecular dynamics study

Candidate

Danilo Di Maio

Advisors

Giuseppe Brancato

Vincenzo Barone

Academic Year 2014-2015

To my family

Contents

The present PhD thesis is based on the following works:

"Breaking the hydrophobicity of the MscL pore: insights into a charge-induced gating mechanism", B. Chandramouli, D. Di Maio, G. Mancini, V. Barone and G. Brancato, *PLoS One*, 10(3), e01201196 (2015).

"Pathways and barriers for ion translocation through the 5-HT3A Receptor Channel", D. Di Maio, B. Chandramouli and G. Brancato, *PLoS One*, 10(10), e0140258 (2015).

"Introducing an artificial photo-switch into a biological pore: a model study of an engineered α-hemolysin", B. Chandramouli, D. Di Maio, G. Mancini and G. Brancato, *Biochim Biophys Acta*, 1858(4), 689-97 (2016).

"Understanding the role of dynamics in the iron sulfur cluster molecular machine" D. Di Maio, B. Chandramouli, R. Yan, G. Brancato and A. Pastore. (Submitted)

Chapter 1

Introduction

Proteins are extremely complex molecular systems whose functions and mutual interplay give rise to various biological events which, as a whole, characterize life. Among the plethora of processes taking place within the cellular environment, one of the most pervasive and, therefore, historically investigated, also due to its ties with extremely influential research fields like neurology and metabolic diseases, is “cell regulation”. Proteins involved in cell regulation fulfil their tasks in many diversified ways, such as directly interacting with other proteins to regulate their functional state, tuning the concentration of ions inside the cell and its compartments, controlling the synthesis of cofactors and key intermediate products within metabolic cycles and so on. Remarkably, advances in our knowledge of these biological macromolecules, from higher and lower organisms alike, turned out to be valuable not only for basic and applied research in biology, but also for seemingly orthogonal fields, such as nanotechnology and analytical chemistry.

Understanding, at molecular level, the way in which such proteins operate implies understanding their microscopic dynamics, which spans several orders of magnitude both in time and length scales. Modern experimental approaches generally provide exceptionally insightful information about protein structural and functional features (like folding, structural transitions, binding and activity), but in many cases their outcomes prove to be cryptic, vague, or controversial, at times. In this context, molecular modelling and simulation demonstrated to be invaluable tools for both rationalizing data and designing new experiments. During the last two decades, impressive advances in theoretical approaches for the investigation

of protein dynamics have been made, owing to a boost in both computational power and accuracy of modelling techniques. These improvements have led to two major consequences. The first one is that the temporal domain accessible to classical simulations extended up to microseconds (μs) for small- to medium-size systems. The second one consists in the possibility to accurately evaluate the energetic landscape underlying the dynamics along selected degrees of freedom, via different free energy calculation methods, thus broadening the spectrum of questions that can be addressed computationally.

This thesis reports on an *in-silico* investigation of a few proteins whose function takes part in cell regulation. In particular, this work focuses on selected proteins that carry out their tasks by modifying either the concentration or the nature of inorganic ions within the cell. Among them, there are systems so versatile to even allow, upon proper biotechnological modifications, the creation of new macromolecular entities exploitable in contexts not strictly biological. According to the growing interest towards these two classes of systems alike, both natural proteins and engineered protein-based nanodevices have been considered within this work.

The thesis is organized as follows. In Chapter 2 a brief introduction to cell regulation is provided. The complexity of cell regulation is mostly described according to a phenomenological approach, in terms of the strategies exploited by cells to perform this activity. The role of inorganic ions within the cell is reviewed along with the most important proteins which take an active part in the regulation mechanisms. In this respect, two pivotal protein classes, i.e. ion channels and pores, are described in detail and the recent advances in protein engineering of these macromolecules are briefly overviewed. Moreover, in view of the critical roles played by inorganic metal clusters within proteins directly involved in metabolic or regulatory pathways, some classes of ubiquitously expressed enzymes, which assemble inorganic ions into such prosthetic groups, are discussed.

In Chapter 3, the computational methodologies used throughout this work are described in some detail. Here, molecular dynamics (MD) simulation techniques are presented along with their advantages and limitations. The physical concept of rare event is presented and the standard approaches used to deal with it are overviewed. A special attention is given to the description of the thermodynamic integration (TI) methods and, in particular, to the Adaptive Biasing Force (ABF) methodology, having been extensively adopted within this work.

The following chapters report the bulk of the original research work carried out during my PhD training program. In Chapter 4, atomistic simulations are exploited to investigate pathways and energetic barriers for ion translocation through a member of the pentameric ligand gated ion channels (pLGIC) family, a well-known superfamily of membrane proteins that mediates communication processes within cellular networks. These protein channels work by converting a chemical signal (represented by an organic molecule commonly referred to as "neurotransmitter") into the alteration of the ionic equilibrium across the cell membrane, through opening of a transmembrane pore. Such alteration brings about an increase in the ionic flux, in the form of an electric spike, which is used to elicit tissue-specific cellular responses. Specifically, the work focuses on one of the most studied members of this family, namely the 5-HT_3A ionotropic receptor, due to its involvement in a wide number of vital neurophysiological functions, and on its peculiar ion transport properties.

In Chapters 5 and 6, I report on the studies of two artificial nanodevices based on engineered bacterial proteins, namely MscL and α-Hemolysin, respectively.

The mechanosensitive channel of large conductance (MscL) is a membrane protein functioning as a release valve in response to osmotic shocks potentially deadly for the cell. It also represents a well-studied model system for the investigation of the molecular bases of mechanosensation, especially for what concerns membrane tension. Due to its favourable transport properties, such as its large conductance and large pore size in the open state, in recent years this system has been increasingly employed to create artificial nanovalves for liposomial vescicles, so as to create nanocarrier prototypes aimed at the controlled release of the most disparate substances, including drugs. In Chapter 5, the artificial gating mechanism introduced within one of such MscL-based nanovalves, so to create a light controllable bio-nanodevice, is examined.

The α-Hemolysin from *Staphylococcus aureus* is a very well-know bacterial pore-forming toxin, which takes its name from its capacity to destroy the integrity of red blood cell plasma membranes, causing their lysis. The capability of α-Hemolysin to form a large, constitutively open and unselective membrane pore has made this protein a particularly appealing scaffold for creating a vast range of different bio-inspired nanopores suitable for analytical chemistry applications. In particular, detection and characterization of analytes is achieved by monitoring the partial current blockages that result from pore obstruction as the analyte translocates through the channel. In Chapter 6 the possibility to create, upon

combining an azobenzene based molecular photoswitch with the *Staphylococcus aureus* α-Hemolysin pore, a light-triggered, controllable nanopore suitable for sensing purposes is investigated.

Chapter 7 reports on a study of the IscS-IscU tetrameric complex, a core protein complex within the complicated machinery of the iron-sulphur protein biogenesis. Specifically, the IscS-IscU complex is appointed with the task of assembling iron-sulphur clusters, a well-known class of ubiquitous inorganic prosthetic groups, and, as such, it is very well-conserved along the evolutionary tree. Recently it has been discovered that, besides its general importance in fundamental biology, this complex (and in particular its malfunction) is also involved a number of human pathologies. Moreover, defects in proteins regulating their functionality by interacting with them during the [Fe-S] cluster biosynthesis stage are also associated with severe diseases. The work presented here is thus aimed at elucidating important steps characterizing the catalytic mechanism of iron-sulphur cluster biosynthesis and the molecular basis of the regulation of the IscS-IscU complex functionality by its allosteric modulators.

Finally, a summary of the main results presented in this thesis and some concluding remarks are given in Chapter 8.

Chapter 2

Biological background

This chapter summarizes the role of inorganic ions in cell regulation and of those proteins which take part to the process. A special attention is reserved to ion channels and protein pores, owing to their important role of allowing the flow of ions, in a controlled or uncontrolled fashion respectively, between the exterior and the interior of the cell, or between cellular compartments, as they will be the subjects of Chapters 4, 5 and 6. The recent advances in exploiting, in non biological contexts and in a controlled fashion, the ion channels and protein pores capability of allowing the flux of solutes between separate compartments, as witnessed by the rise of bio-nanotechnology, is overviewed. Besides, the structural and functional role of inorganic clusters within regulatory proteins, and therefore the importance of the metabolic pathways in which they are synthetized, are discussed, as introduction to Chapter 7.

2.1 Cell regulation and homeostasis

Cell regulation is a broad term, usually encompassing the many integrated processes that occur within a cell and that, as a whole, are aimed at maintaining homeostasis. The concept of homeostasis[1] refers to a balanced condition in which cells (and on a larger scale tissues and organisms) keep their physico-chemical conditions stable and relatively constant, avoiding harmful extremes of any form, through various regulatory processes, both active or passive (a general

mechanism is sketched in Figure 2.1). Regulatory processes moderate everything, from the growth and replication rates of cells, to the proteins and metabolites concentration, to the acidity and ionic strength levels of the cellular environment. In a physiological situation, these crucial variables are allowed to vary,[2] for instance in response to external stimuli or to a change in the cell internal environment, but always within tight limits. Such limits are needed to maintain the complex set of interacting metabolic reactions all organisms depend on. If a breakdown in the regulatory mechanisms occurs and an homeostatic imbalance is reached, detrimental events may take place,[3] with, if persisting, potentially pernicious, or even deadly, consequences for the organism.

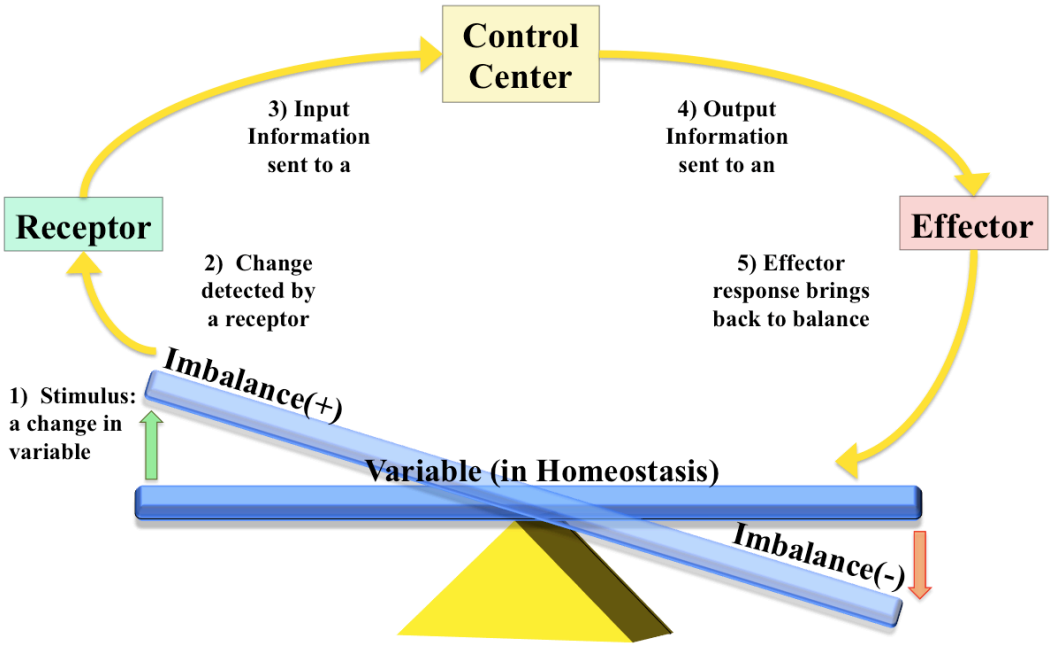


Figure 2.1: Schematic representation depicting the main steps of the cell regulation mechanism responsible for the control of homeostasis. The three key actors in this mechanism, here enclosed in coloured boxes, may be located in different cellular compartments or, in the case of more complex organisms, even in different cells or tissues.

2.1.1 Inorganic ions in cell regulation

Inorganic ions constitute an essential component for vital cellular activity in all domains of life.[4,5] Their relevance is supported by an estimate according to which metal ions are essential for the function of over one third of all proteins and of almost half of all enzymes.[6] A few of them are present as trace elements, especially transition metals, and usually coordinated to proteins, while others are more abundant, such as Na^+, K^+, Ca^{2+}, Cl^-, phosphate and bicarbonate and usually found as free solutes.

The biological role of inorganic ions goes far beyond the simple oxygen transport, as in haemoglobin, enzymatic catalysis, as cofactors within metalloenzymes, or electron transfer.[7] Indeed, they are necessary for many of the most important biochemical processes, i.e. protein and nucleic acid folding, protein aggregation, allosteric regulation, cell signalling, cell adhesion, cell-cell recognition, transcription regulation and biomineralization. Transition metal ions, above all Zn^{2+} but also Fe^{2+} and Fe^{3+},[8,9] in the form of iron-sulphur clusters, are found within structural motives in transcription factors, nucleases and chaperons, mainly coordinated by cysteine or histidine residues. Alkaline earth metal ions, such as Mg^{2+} and Ca^{2+}, are present in cadherins and integrins,[10,11] two membrane protein families involved in cell-cell and cell-extracellular matrix adhesion processes. Calcium ions also play a pivotal role in many signal transduction pathways,[12] where they act as a second messenger, either directly or bound to Ca^{2+}-modulated proteins like calmodulin. Last but not least, alkali metal ions, mainly Na^+ and K^+, along with chloride, phosphate and bicarbonate anions are used by transporters to set the osmotic pressure and pH equilibrium between the cell and the external medium or between cellular compartments.[13] Besides, ion pumps, like the Na^+K^+ ATPase, use the ATP energy to concentrate Na^+, Cl^- and Ca^{2+} and K^+, on different sides of the plasma membrane (or within storage sites). Normally, the concentration of Na^+, Ca^{2+} and Cl^- is higher outside the cell (even though Ca^{2+} is abundant in mitochondria), while that of K^+ is higher inside the cell, making, overall, the cell interior and the cell exterior negatively and positively charged, respectively.[14] The membrane is then said to be *polarized*. This charge polarization gives rise to an electrical potential difference, the so called *membrane potential*, which, along with the ionic concentration differences between both sides, creates an electrochemical gradient across the membrane. Such gradient is particularly important, since it provides the driving

force for many biological events at the membrane level, among which stands out electric signalling.

2.2 Inorganic ions and electric signalling

Electric signalling is one of the cell signalling mechanisms, i.e. the set of processes through which cells are able to sense and adequately response to the state of their microenvironment or to an information forwarded by other cells.[15] Electric signalling takes place in cells which are endowed with a particular property: they are "*excitable*". This means that they are able to detect an external signal and convert it into an electrical signal (specifically a change in membrane potential), which, in turn, can either be passed on to other cells or trigger an appropriate biological response. Pivotal processes, like neuronal communication, muscle contraction, heart beating, or chemical transmitters secretion, depend on such signals.[16]

The electric signal is generated by the flow of ions across the membrane and the change in membrane potential it brings about. Inorganic ions tend to move across the membrane following a particular direction, which is dictated by their electro-chemical gradient, given by the following formula.

$$\Delta G = RT \ln([C]_{in}/[C]_{out}) + Q\mathcal{F}V_m$$

The gradient ΔG that causes an ion (e.g. a cation) to pass inward through the membrane is a function of the ratio of its concentrations between the two sides of the membrane ($[C]_{in}/[C]_{out}$) and of the membrane potential (V_m), with R being the gas constant, T the absolute temperature, Q the charge of the ion and $\mathcal{F}$ the Faraday constant. The membrane potential can be estimated through the Goldman-Hodgkin-Katz (GHK) voltage equation,[17] which, in the case of monovalent species only, has the following form

$$V_m = \frac{RT}{\mathcal{F}} \ln\left(\frac{\sum_i^N P_{M_i^+} [M_i^+]_{out} + \sum_j^M P_{A_j^-} [A_j^-]_{in}}{\sum_i^N P_{M_i^+} [M_i^+]_{in} + \sum_j^M P_{A_j^-} [A_j^-]_{out}}\right),$$

$[M_i^+]_{out}$, $[M_i^+]_{in}$, $[A_j^-]_{out}$ and $[A_j^-]_{in}$ being the concentrations of monovalent cations (M) and anions (A) outside or inside the cell, respectively, and $P_{M_i^+}$ and $P_{A_j^-}$ being the relative permeability of the membrane to such ions. In this specific formulation Ca^{2+} is omitted, but it can be added to deal with situations in which it plays a significant role. Note that, since the charges of cations and anions have opposite sign, their concentration ratios are reversed. So, in simple terms, according to the GHK model the membrane potential is a function of the external/internal concentration ratios of each ion, weighted by that ion permeability. Thus, the resting membrane potential basically results from the ion permeability that predominates when the cell is at rest. For instance, Na^+ has almost no resting permeability and a very high driving force (ΔG) to move inward, while K^+ permeability is very high and its driving force is very low. For this reason, at rest, the V_m in excitable cells (-60 mV) is closer to the equilibrium potential for K^+. Figure 2.2 provides the sketch of an excitable cell at rest.

A given ionic species flows across the membrane as long as the combination of concentration gradient and electrical potential provides a driving force. The number of ions that must flow to produce a physiologically significant change in the membrane potential is negligible relative to the concentrations of Na^+, K^+ and Cl^- in cells and extracellular fluid. Therefore, the ionic flux that occurs during signalling in excitable cells has essentially no effect on the concentration of these ions. With Ca^{2+} the situation is different; because the baseline intracellular [Ca^{2+}] is generally very low, ~10^{-7} M (as mentioned above, Ca^{2+} is stored in mitochondria), inward flow of Ca^{2+} can significantly alter the cytosolic [Ca^{2+}]. The ionic flux also causes a redistribution of charges on the two sides of the membrane, changing V_m. The net influx of a positively charged ion, such as Na^+, or efflux of a negatively charged ion, such as Cl^-, *depolarizes* the membrane and brings V_m close to zero. Conversely, the net efflux of K^+ *hyperpolarizes* the membrane and V_m becomes more negative. At the equilibrium potential (E), i.e.

$$E = -\frac{RT}{Q\mathcal{F}}\ln([C]_{in}/[C]_{out}),$$

the driving force (ΔG) tending to move an ion is zero. The equilibrium potential is different for each ionic species, according to the aforementioned formula, because the concentration gradients differ.

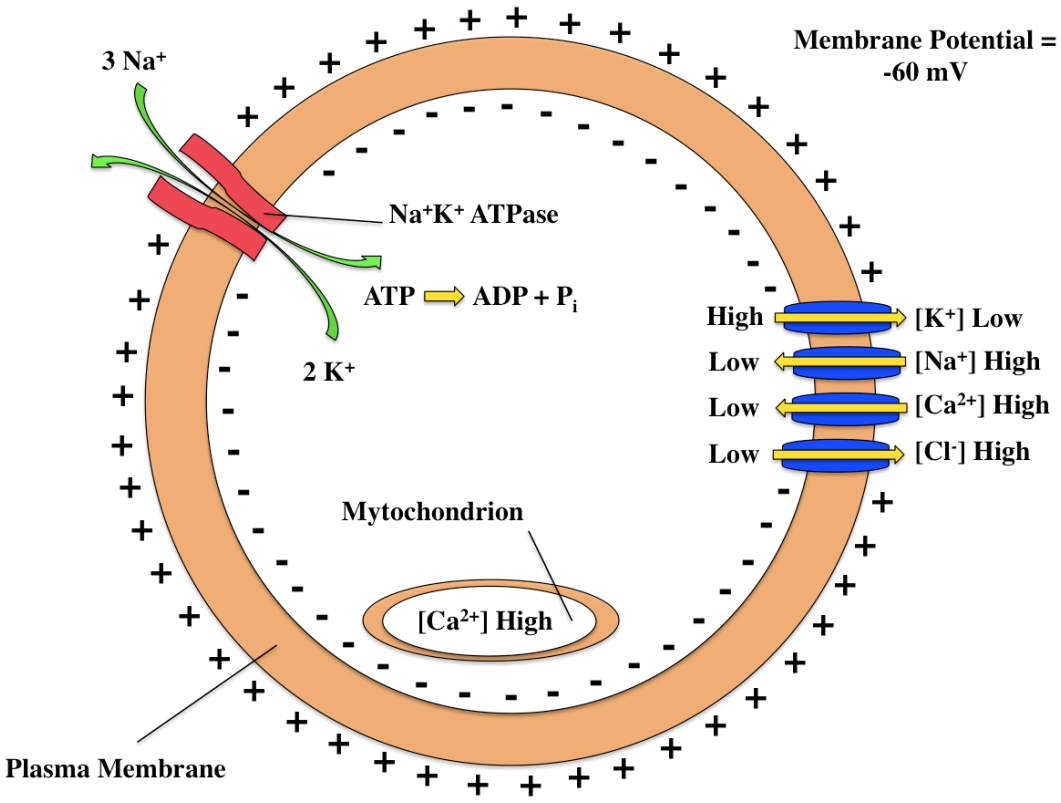


Figure 2.2: Schematic representation of the electrical signalling mechanism. The electrogenic pump Na^+K^+ ATPase (displayed in red) generates a transmembrane electric potential of about -60 mV (inside negative). The yellow arrows show the direction in which ions tend to move spontaneously across the plasma membrane in animal cells, driven by the combination of chemical and electrical gradients.

Proper cellular function requires membrane permeability to a particular ion to be high at certain times, but not at other times, therefore leading to a variable ionic conductance. The lipid bilayer, however, is highly resistant to the movement of electrically charged species, as it functions mainly as an insulator. For this reason, cell membranes incorporate proteins, known as *gated ion channels*,[18] responsible for creating paths through which ions can flow inward or outward.

The term “gated” refers to the property of being opened and closed in a controlled fashion, specifically in response to stimuli of chemical or physical nature. Different cells express various types of gated ion channels, which differ for the specific signal sensed and the selectivity toward specific ions (or group of ions). The latter is usually achieved by winnowing ions out according to their charge and/or ionic radius. The membrane potential of a cell at a given instant is ultimately the result of the precisely timed gating of different types and numbers of these channels. When a channel opens, the consequent increase in membrane permeability for that particular ion (or group of ions) shifts the membrane potential towards the membrane potential for that ion. This may lead to the opening of other ion channels or to the activation of intracellular pathways. Once the signal ceases, ionic pumps come into play to restore the rest membrane potential.

The regulated and transient ionic flux realized by gated ion channels is essential to bring about a biological response of controlled degree and limited in time (usually in the millisecond range).[19,20] Indeed, the loss of control on the plasma membrane permeability to ions, and, as a consequence, a breakdown in the upkeep of the electro-osmotic gradient, would result in the dysregulation of all the downstream cellular processes. Detrimental consequences can ensue, including several pernicious disorders, with a potentially widespread damage for the entire organism.[21,22] That being said, it may not be surprising that many of the most potent microbial toxins operate by compromising the plasma membrane integrity or permeability, and, therefore, also the electrical signalling process and cell homeostasis. In particular, within the broad category of the membrane damaging toxins (also known as cytolysins) a peculiar class of proteins, known as *pore-forming toxins*,[23] stands out. The toxicity of such pore-forming proteins arises from their ability to create open, unregulated pores within the lipid bilayer, thereby promoting water influx and dispelling the electro-osmotic gradient. In many cases, the size of the pore can be so large that even molecules greater than simple ions (from ATP up to small proteins) may be allowed to cross the plasma membrane,[24] eventually leading, in susceptible cells, to death.

Hence, given the importance electric signalling regulation or dysregulation have on vital cellular functions, understanding, from a structural point of view, how gated ion channels and pore-forming proteins operate is momentous. Such knowledge can prove useful in several ways. For instance, it can provide the grounds for a pharmacological intervention in case of gated ion channel over- or

under-stimulation; or, it can be used for aptly modifying their structures so as to alter their transport properties or introduce completely new features. Accordingly, in anticipation of Chapter 4, 5 and 6, the classification and the most important structural and functional features of gated ion channels and pore-forming proteins will be described in the next paragraphs.

2.3 Gated ion channels

Gated ion channels (GICs) were first recognized as independent membrane inserted molecular entities back in the late 1960s, during electrophysiological studies on nerve and muscle cells.[25] Afterwards, thanks to the boost in genomics, the ancient origin of this protein class in the evolutionary timeline became clear, as its members are shared by almost all life forms. The GIC protein class encompasses a huge number of members, whose primary role is to permit that ions cross the membrane at certain times, but not at other times, in response to an array of diverse physico-chemical stimuli. Such stimuli include: changes in the membrane potential (sensed by voltage-gated ion channels);[20] osmotic stress and mechanical vibrations (sensed by mechanosensitive channels, MSCs);[26] temperature changes;[27] pH changes;[28] light;[29] binding of ions and metabolites (sensed by ligand-gated ion channels, LGICs). [30]

Most of our knowledge about the GICs structural organization comes from very recent X-ray crystallography and cryo-electron microscopy (Cryo-EM) studies.[31-33] Almost all GICs are integral oligomeric membrane protein complexes (with few exceptions) formed by the assembly of identical or related subunits, ranging from two up to seven, around an axis of rotational symmetry. The single subunits are separately synthetized (and, if needed, modified in the Golgi's apparatus through, e.g., N-glycosylation) and subsequently directed to the membrane surface. Once there, alone or assisted by chaperons, they embed into the membrane and assemble to form the channel.

GICs are characterized by structural topologies and quaternary structures very different from each other (though proteins belonging to the same superfamily usually share similar structures) and can be formed by one or more domains (as Figure 2.3 shows). Despite such inter-family diversity, they all share a common structural feature: the presence of a transmembrane (TM) domain, made up by a bundle of alpha-helices, framing a central, funnel-shaped pore which provides a

permeation pathway through which ions diffuse, usually very fast, along their concentration gradients.

The channel gate resides in the TM domain. At this level, frequently corresponding to the pore maximum constriction point, the channel is lined by a stretch of non-polar residues, which make up a sort of "hydrophobic girdle", hindering the passage of water or ions.[34,35] Specifically, it occurs that, in correspondence of this critical region, the combination of hydrophobic effect and small pore diameter makes the behaviour of water and ions in the channel interior very different from the situation in the bulk. Indeed, as also observed in model nanopores, there exists a diameter threshold value, estimated around 9 Å, below which water density resembles more that of the vapour state rather than that of the liquid state.[36,37] This is reflected by a partial (or even total) dewetting of the pore surface, which in turn creates an highly effective energetic barrier to ion or even water permeation, making, under resting conditions, the channel gate closed. Over the last decade, there has been a growing number of experimental evidences in support of a gating mechanism being based on the hydrophobic effect and shared by all GICs.[38-40]

The opening of the gate is the result of an allosteric process involving a complex conformational rearrangement.[41] Such rearrangement originates at the protein site in direct contact with the environmental signal and propagates throughout the protein structure, eventually ending to the transmembrane pore. At present, it is commonly accepted that the gate opening involves an increase in the diameter at the pore narrowest region, so as to shift the equilibrium between wetting and dewetting of the pore in favour of the wet state. However, the gating process itself is still, in large part, not completely understood, since it takes place on a very long temporal (and for some channels also spatial) scale.

Additionally to their structural diversities, GICs also display functional differences, usually related to the previous ones, in their transport properties, such as ionic selectivity and conductance. For instance, some GICs are specific for a single ionic species (e.g. Na^+ or K^+), some other are selective for a particular group of ions (e.g. cations or anions) and some can be even completely non-selective. Ionic selectivity is achieved through the presence of structural elements, hosted in the TM domain, known as selectivity filters. From a structural point of view, these filters vary considerably among GICs, as they differ in length, width, amino acid composition and charge (they can be neutral or charged) and can be made up by backbone or side chains atoms.

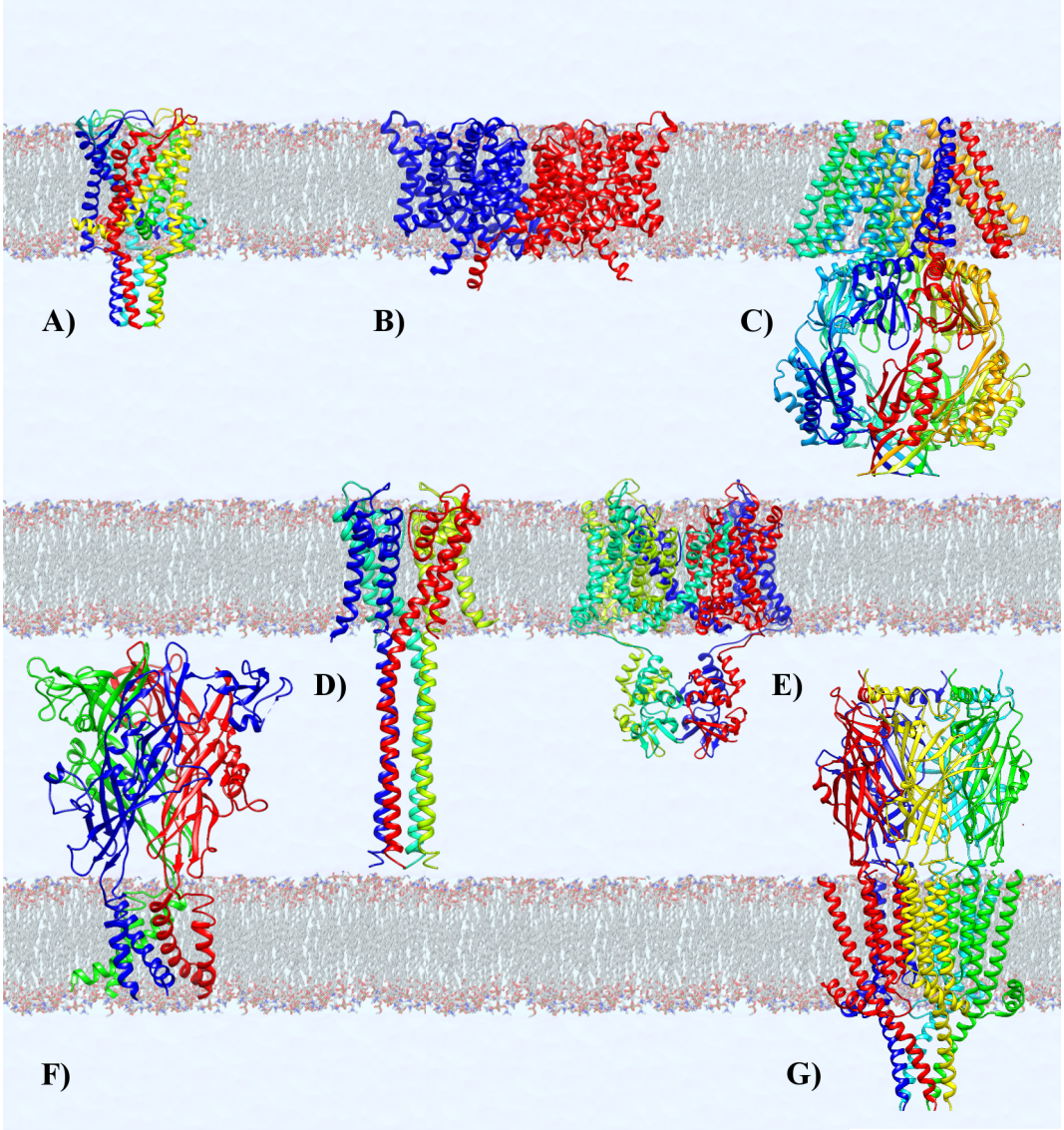


Figure 2.3: Few examples of gated ion channels three-dimensional structures embedded within a lipid membrane. A) Mechanosensitive channel of Large conductance (MscL) (PDB code: 2OAR);[42] B) bacterial chloride channel (ClC) (PDB code: 1KPL);[43] C) Mechanosensitive channel of Small conductance (MscS) (PDB code: 2OAU);[44] D) proton activated K^+ channel (KcsA) (PDB code: 3EFF);[45] E) voltage gated K^+ channel (Kv) (PDB code: 2R9R);[46] F) zebrafish P2X4 receptor (PDB code: 3H9V);[47] G) 5-Hydroxytriptamine receptor of type 3 (5-HT_3R) (PDB code: 4PIR).[48]

The structure of the selectivity filter can have a profound impact on the solvation state of the ions, as it can either allow them to translocate with an intact first hydration shell or compel to a partial dehydration in order to move through it. Altogether these factors contribute to determine the energetic landscape experienced by ions inside the filter and, therefore, the selectivity towards a particular ion or group of ions.

The ionic conductance, i.e. the permeation rate, is often correlated with the channel dimension and can range from very low values, like few pS, for narrow TM channels, as in the case of pentameric Ligand-Gated Ion Channels (pLGICs),[49] to high values, in the order of nS, as in the case of larger channels like the Mechanosensitive Channel of Large Conductance (MscL).[50] However, it has to be pointed out that, at variance with ionic selectivity, channel conductance is usually a more "global" property. That means it can be influenced by structural elements distributed all over the channel surface, hence possibly present on further domains and not strictly localized in the TM domain. As a result, ion conductance can be quite different also among channels of the same family.

Understanding how GICs regulate the amount and the charge of translocating ions, thereby tuning the ionic flux, is pivotal to understand cellular communication, especially at the chemical synapse level. In order to get insights into the transport properties of a GIC member, namely the 5-Hydroxytryptamine receptor of type 3 (5-HT_3R), we carried out an thorough molecular simulation study, which is reported in Chapter 4.

2.4 Pore forming toxins

The concept of membrane pore was first elaborated to explain the lytic damage brought about by the complement system to bacteria or other non-self cells.[23] However, sooner it was realized it could extend to a broader range of lytic agents.[51] In 1979, the classifications of cytolysins of different origins clearly indicated that many of these proteins don't cause cell death purely by membrane destruction but rather form discrete lesions (i.e. pores) in the plasma membrane.[52] Since then, the pore forming toxins (PFTs) have been identified as one of the major and widespread weapons for chemical warfare in a wide range of organisms, both prokaryotic and eukaryotic, including bacterial, fungi, plants and

even animals. However for practical reasons, as well as the importance for their disease-causing abilities, the pore-forming toxins of bacteria have received far more attention than those from any other kind of organism. Nonetheless, due to sequence homology conserved across vast evolutionary distances, in few instances PFTs show structural and mechanistic similarities, thereby allowing conclusions drawn from the investigation of the bacterial orthologues to be generalized to proteins belonging to other organisms.

The PFT protein class encompasses several members, which are characterized by many structural and functional features that make them fundamentally different from GICs. First and foremost, they are not gated by any environmental or intracellular stimuli but instead they create constitutively open, water filled pores (though exceptions exist), through which ions and bigger solutes can freely diffuse. Such free diffusion dissipates the electro-osmotic gradient and lead to cell death by loss of electrolytes and key organic metabolites like ATP. Cell membrane damage may have important fallouts besides the lytic effect per se. For instance, it can facilitates the uptake of essential nutrients from damaged cells, or the bacterial spread trough the tissues of the host, in concert with enzymes released by the invading bacteria which degrade cellular matrix.[53,54]

Secondly, PFTs are dimorphic[23] since they exist in two structurally distinct forms, that is as water soluble protein monomers, suitable for diffusion in the organic fluids, or as membrane inserted protein complexes, the latter being the effectively harmful form of the toxin. The insertion process, conversely to what happens for many GICs where the membrane insertion requires the assistance of a number of proteins, is completely spontaneous and usually involves the homo-oligomerization of several protein monomers. The pore forming mechanism is comprised by three fundamental steps. The first step involves the synthesis of the water soluble citolysin monomers and their release in the extracellular medium. The second step involves the PFT monomer adhesion to target cell membrane, a process which is apparently achieved through the interaction with specific membrane components, usually lipids. Subsequently, the PFT monomers diffuse laterally in the membrane until they encounter other membrane bound toxin monomers. Once this happens, the PFT monomers, via specific protein-protein interactions (PPIs), rapidly oligomerize to yield an intermediate homomeric complex termed as “pre-pore”. Such oligomer is characterized by a relatively specific size and represents a pre-insertion state where the monomers interact so strongly to make it resistant to dissociation by many detergents and chaotropic

agents. Ultimately, in the third stage the PFT oligomer penetrates the target cell membrane with its TM domain, forming the final functional membrane pore. The oligomeric pores can differ widely in size. The smallest pores formed in mammalian cell membranes have been sized to around 1-2 nm effective diameter.[55] In contrast, the TM pores formed by cholesterol binding cytolysins (CBCs), a sub-class of PFTs, can consist of up to ~ 50 monomers and create holes up to 25-30 nm diameter.[56]

The progress in the cloning and sequencing of PFT structural genes and in the knowledge of their mechanism of action allowed to pinpoint another important difference with GICs, i.e. that PFTs, though sharing the mechanism of action, often don't form known families. Specifically, at first level two PFT categories, present in both Gram-positive and Gram-negative bacteria, can be distinguished: "cogeneric" and individualistic" toxins.[57] The first category comprises those proteins that share close structural similarities and/or similar physicochemical and functional lytic properties and mechanisms of action, such as CBSs, pore-forming RTX (repeat in toxins), bicomponent toxins from *Staphylococcus*, *Serratia* and *Proteus* and some hemolysins from *Vibrio* species. The latter category, instead, comprises toxins exhibiting and appreciable degree of uniqueness, such as *Staphylococcus aureus* hemolysin and *Aeromonas* aerolysin.

Several biophysical studies have been carried out on both soluble monomers and assembled pores to in order to characterize and classify their structure.[23] Despite the structural knowledge of the monomeric form, as compared with that of the assembled pore, can provide significant clues about the mechanism of pore formation and membrane insertion, the toxin structural classification through monomer structure is not very convenient. Indeed, the water soluble folds they adopt are just temporary and often quite different from those in the assembled state, which makes the structural classification of the holigomers, the effectively active form of the toxin, more informative. From a structural viewpoint, PFT oligomers can be roughly assigned to two categories, according to the predominant secondary structure element of the pore in the TM region (as shown in Figure 2.4): those formed predominantly by α-helices (α-PFTs) and the β-sheet rich PFTs (β-PFTs), which constitute the broader category.

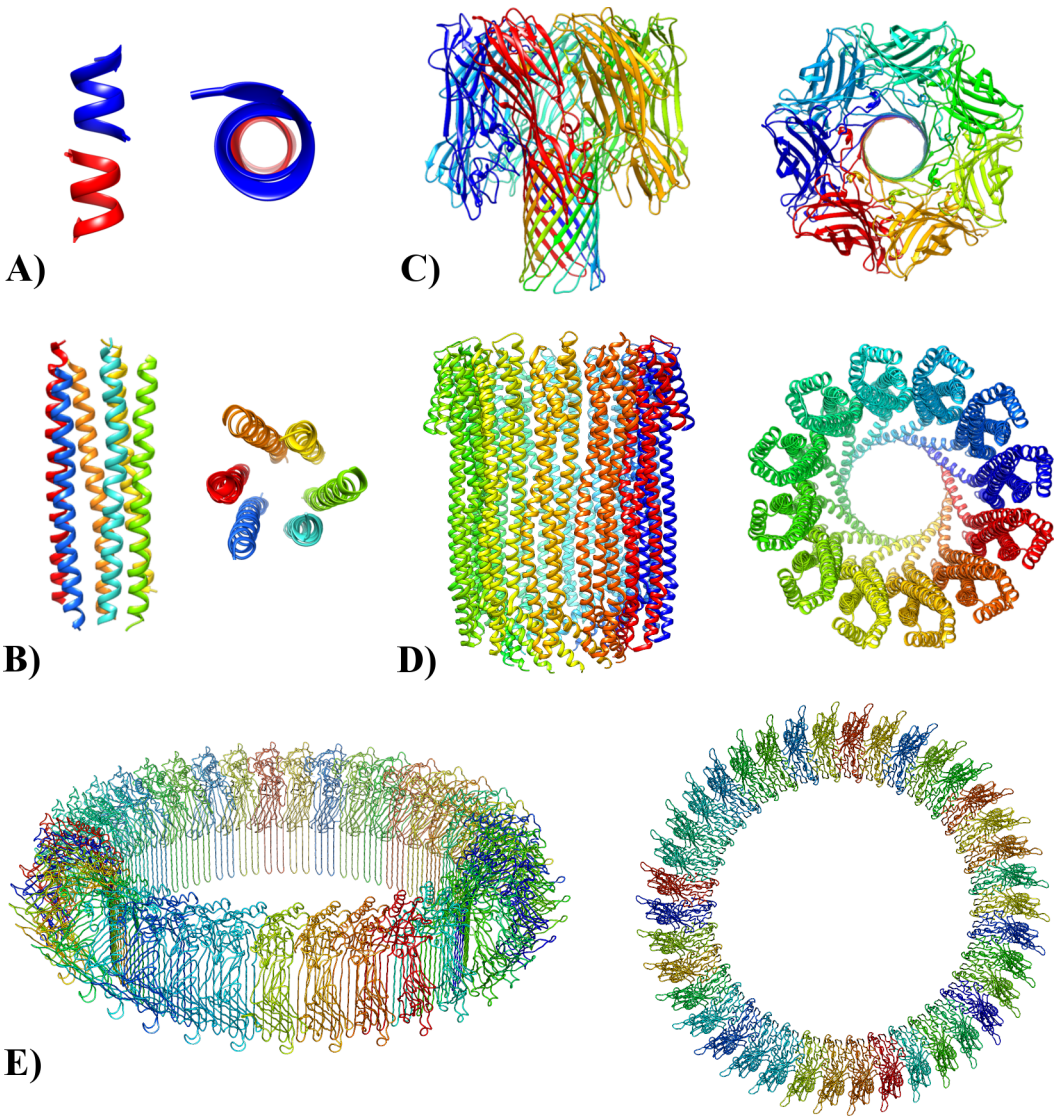


Figure 2.4: Few examples of pore forming toxin three-dimensional structures. A) Gramicidin A (PDB code: 1JNO);[58] B) Dermcidin (PDB code: 2YMK);[59] C) *Staphylococcus aureus* α-hemolysin (PDB code: 7AHL);[60] D) Cytolysin A (PDB code:2WCD);[61] E) Cryo-EM model of the assembled Pneumolysin pore (PDB code: 2BK1)[62]. Both side view (left) and top view (right) are displayed.

The pore of α-PFTs can be made up by one single α-helix or a bundle of α-helices per monomer, arranged around an axis of rotational symmetry. The TM

helices tend to be very hydrophobic and adopt the same conformation in both assembled and water-soluble state, where are hidden within the fold of the protein. Conversely, the pore of β-PFTs is a β-barrel made up by several amphipathic β–hairpins (at least one per monomer), which expose the hydrophobic side to the lipids and the hydrophilic side to the pore interior. The TM regions of β-PFTs adopt different folds in the water soluble states and are converted to β–hairpins only upon membrane insertion. A somehow peculiar group is that of pore forming antimicrobial peptides (PFAPs), which includes several relatively small linear or cyclic amphipathic peptides, formed by usually ~12 to 50 L- or D-amino acids (or a mixture of both),[63] predominantly hydrophobic (usually ~ 50% of the sequence) and basic.[64] PFAPs are normally too small to create membrane holes and thus, after insertion, oligomerize and form bigger proteinaceous[65] or protein-lipid pores,[66] featuring a variety of secondary structures, from α-helices to β–sheets to a combination of both.

2.5 Engineered protein channels and nanopores

GICs and PFTs are a topic of interest not only for physicians or microbiologists but also for a variety of other scientists, from biochemists to nanoscientists. In recent years indeed, engineered biological channels and pores have been fruitfully used for various biotechnological applications.[67-69] Specifically, nanopore engineering was successfully used to enhance or add new functionalities to such systems, e.g. to regulate molecular transport by pre-determined external stimuli.[70] Efforts focused on introducing artificial triggers and switches to control and modulate the nanopore functionalities have shown their utility in versatile applications as diverse as drug delivery, molecular sensing, disease detection and reaction monitoring.[68,71] One of such engineering strategies concerns the use of photochromic ligands, capable of changing their structure and properties in response to light. The strategy of tethering photo-switchable moieties into biomolecules has provided an unprecedented temporal control of purposely designed nanodevices,[72] as demonstrated, for example, by the light-mediated regulation of the activity of enzymes and biological channels. Already, a number of studies have reported the photomodulation of biological activity using tethered photochromic ligands. Some pioneering examples include I) the reversible photoactivation of nAChR receptor,

II) the dimerization of gramicidin-A by a photoswitchable cross linker, III) the photocontrol of K+ channels based on the photoisomerization of azocompound, IV) the activation of iGluR receptor by photoswitch ligated agonist, V) the modulation of *Escherichia coli* MscL activity via covalent linking of photo-caged compounds, VI) the blocking of hemolysin pore lumen via conjugating sulfonated azobenzene.[73-77] Among the various possible applications of these devices there is the fast-growing field of nanopore analytics, i.e. single-molecule detection using nanopores. Such an intriguing idea relies on correlating the transient current blockades induced by the analyte, as it passes through or binds the pore, to its molecular properties, such as size, charge or folding.[71,78] To this aim, non-specific biological pores are functionalized, using site-specific mutagenesis and synthetic chemistry, in order to introduce an artificial gating and/or selective blockage. Moreover, introducing such a new functionality through a light-activated molecular device may provide a fine temporal control of the pore gating, without the need to resort to more invasive techniques for the environment, as in the case of pH- or redox-based triggers.

Modelling and simulation of such nanopore based systems can provide useful insights into their functioning mechanism as well as serve as a valuable tool for the design of novel bio-inspired molecular devices. The computational study of two artificial nanopores based on a GIC (the MscL channel) and on a PFT (the *Staphylococcus aureus* α-hemolysin) is reported, respectively, in Chapter 5 and 6.

2.6 Iron homeostasis and the [Fe-S] cluster biogenesis machinery

Iron is essential for life[79] being the most abundant transition metal in human body and one of the most biologically relevant inorganic ions. Iron ions are involved in essential functions of life, ranging from redox chemistry to oxygen transport and even cell regulation, owing to its versatile chemical and electronic features. Iron is present only in extremely small trace as a free solute, at variance with other more abundant ions in biological fluids, like Na^+, K^+ and Cl^-. Indeed, whereas free Fe^{3+} is relatively inert and poorly soluble at pH 7, free Fe^{2+} is highly reactive and extremely cytotoxic. Specifically, Fe^{2+} is able to directly catalyse, via Fenton chemistry,[80,81] the formation of damaging reactive oxygen species (ROS), such as superoxides, peroxides and free radicals, which can cause severe

oxidative damage to cell constituents like lipids, proteins and DNA, and thus are extremely harmful to the organism.[82]

Iron homeostasis, including uptake, metabolism and storage, is thus very tightly regulated and every step in all the aforementioned processes is associated with a redox cycling of iron oxidations and reductions between $Fe^{2+}$and Fe^{3+}.[83] The latter is the dominant form of iron we assume from diet and is conveyed to other cells through bloodstream circulating transferrin.[84] Subsequently, iron delivery into the cell interior takes place through receptor-mediated endocytosis of plasmatic iron loaded transferrin (Tf), upon interaction with the transferrin receptor (Tf-R).[85] Such mechanism is, however, both organism and cell-type specific. Once inside the cell iron fate is dual: it can either be dispatched to storage sites or it can be channelled into the enzymatic machineries that are responsible for iron incorporation into its biologically active forms, e.g. cofactors. Cellular iron is stored into ferritin,[86] a big globular protein complex of 450 kDa, consisting of 24 protein subunits, and represent the primary intracellular iron-storage protein in both prokaryotes and eukaryotes, being able to host about 4500 Fe^{3+} ions. Ferritin is endowed with a ferroxidase activity, which allows it to convert iron from ferrous to ferric state, so as to store it as inert, non toxic ferrihydrite mineral. When necessary, iron reserves can undergo reduction-induced mobilization by electron donors, resulting in the release of Fe^{2+} ions, ready to be transformed according to the needs of the cell.

Concerning the biologically active forms of iron, one of the most common is certainly that of inorganic iron-sulphur ([Fe-S]) clusters,[87,88] which are often found within structural motifs of regulatory proteins, where they play important structural and regulatory roles. Since the discovery of ferredoxins in the early 1960s,[89] the number of identified proteins containing [Fe-S] clusters, referred to as [Fe-S] proteins, has greatly increased. The discovery of many different types of [Fe-S] proteins and [Fe-S] clusters has resulted in an appreciation of their remarkable functional and structural diversity (as shown in Figure 2.5).

[Fe-S] clusters can play purely structural roles, as in the case of the DNA repair enzymes endonuclease III[90] and MutY [91], where evidences indicate that redox-inactive cubane type [4Fe-4S] cluster, similarly to Zn in Zn-finger proteins, controls the structure of a protein loop essential for recognition and repair of damaged DNA. Moreover, there are now several well-characterized examples of [Fe-S] proteins involved in transcriptional or translational regulation of gene expression in bacteria.[92] In these cases, each [Fe-S] cluster senses a

different type of environmental stimuli and uses a distinct sensing mechanism, involving cluster assembly, conversion, or redox chemistry, in order to promote or repress the synthesis of proteins responsible for eliciting a specific biological response. For instance, the SoxR protein senses oxidative stress via oxidation of the $[2Fe\text{-}2S]^{2+}$ cluster and thereby stimulates transcriptional expression of SoxS, which is responsible for activating the transcription of numerous enzymes in the oxidative stress response.[93]

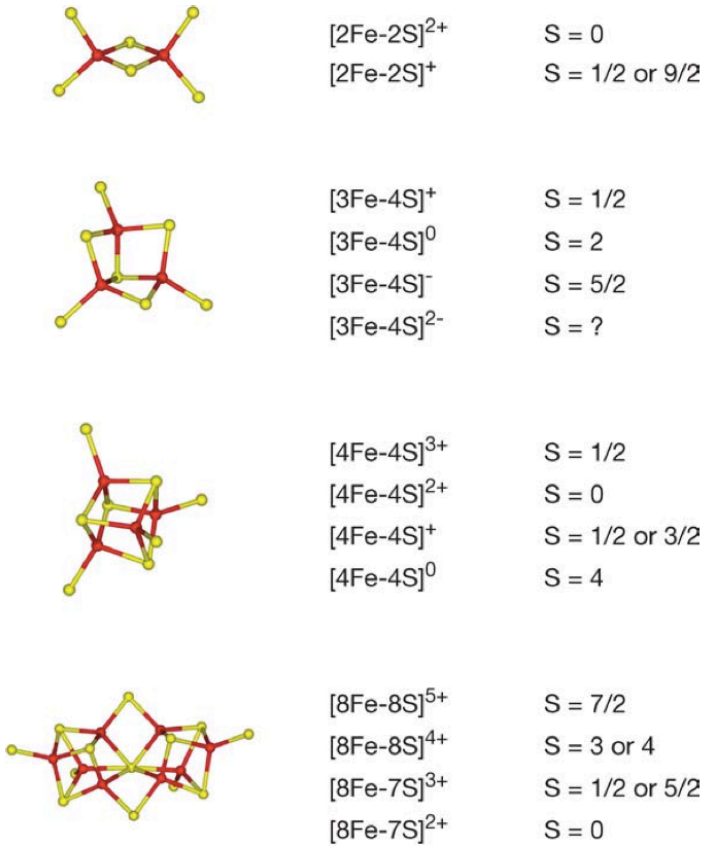


Figure 2.5: Structures, core oxidation states, and spin states of crystallographically defined [Fe-S] clusters. Iron is shown in red, and sulphur is shown in yellow. The spin state, denoted by a question mark, has yet to be determined.

The FNR (fumarate and nitrate reduction) protein is an oxygen sensing transcriptional regulator that functions by oxygen-induced conversion of the DNA-binding $[4Fe\text{-}4S]^{2+}$ cluster to $[2Fe\text{-}2S]^{2+}$ cluster, so as to control the

expression of genes involved in the aerobic and anaerobic respiratory pathways of *E. coli*.[92] Finally, IRP1 (iron-regulatory protein 1), which binds to and stabilizes mRNA iron-responsive elements in eukaryotes and consequently regulates the translation of ferritin and transferrin receptors, is controlled by the assembly or the loss of a [4Fe-4S] cluster on the same IRP.[94]

Given the broad spectrum of functions carried out by [Fe-S] proteins, it should not be surprising mutations in genes that encode proteins involved in human [Fe-S] protein biogenesis may have pleiotropic consequences. Indeed, in the past two decades, several diseases have been attributed to defective [Fe-S] protein biogenesis systems.[95] Consequently, understanding the actors and the mechanism of [Fe-S] protein biogenesis is of paramount importance in medicine.
Most of our knowledge about the eukaryotic [Fe-S] protein biogenesis derives from studies in bacteria and yeast.[96,97] The whole [Fe-S] protein biogenesis process can be generally divided into two fundamental steps: 1) formation of the [Fe-S] cluster and 2) cluster insertion into apo-proteins. It is now widely recognized that the [Fe-S] protein biogenesis machinery is not unique but rather many different types exist,[88] which usually require a complicate interplay of a large number of proteins. These different and often interconnected[98] machineries can be located either in mitochondria or in cytosol and can be charged with specialized functions, such as providing [Fe-S] clusters for the maturation of specific proteins,[99] e.g. nitrogenases, or activating only under particular metabolic/environmental conditions, e.g. oxidative stress or iron starvation.[100] However, despite their specificities, all of them are mechanistically unified by the requirement for two proteins: a cysteine desulphurase, which provides the elemental sulphur, in the form of a highly reactive persulphide carried by an active site cysteine, for the cluster assembly and an [Fe-S] cluster scaffold-protein, on which the cluster is assembled and subsequently transferred to acceptor proteins.

While the overall [Fe-S] protein biogenesis mechanism of these different machineries has been known for a number of years, the detailed molecular mechanism for each step remains to be elucidated. Indeed, there are many intriguing mechanistic questions related to the [Fe-S] protein biogenesis process that are still poorly understood, such as how [Fe-S] clusters are transferred from scaffolds to target proteins, how various accessory proteins participate in [Fe-S] protein maturation, how the [Fe-S] cluster biosynthetic process is regulated and, ultimately, how [Fe-S] clusters are assembled on scaffold proteins. Among these,

the latter is the most extensively addressed, but at the same time less resolved, due to the intrinsic biochemical complexity of the [Fe-S] cluster synthesis stage. Thus, to shed some light on the molecular details of a few key steps of the [Fe-S] cluster assembly mechanism, we have carried out a molecular simulation study of the *E.coli* IscS-IscU complex, i.e. the desulphurase/scaffold-protein pair of one of the most investigated [Fe-S] protein biogenesis machineries, namely the ISC (Iron Sulphur Cluster) system.[101] The results of our simulations are reported in chapter 7.

Chapter 3

Theoretical and computational methods

Biological macromolecules typically display a wide range of characteristic motions at different time and space scales, ranging from fast, localized motions of small chemical groups, to slow, large-amplitude, collective motions involved in folding transition or domain swapping. Molecular dynamics (MD) techniques are, in principle, capable of following atomistically the structural rearrangements of a molecular system as a function of time and to relate the outcome of molecular simulations to relevant structural, dynamic and thermodynamic properties. Indeed, MD has been successfully applied to the study of protein stability, conformational transitions, ionic transport and interaction with molecular partners. As a result, in the last decades MD simulations have become a wide-spread and well-established tool to understand the mechanism underlying complex biological processes, e.g. for interpreting experimental data or for making novel predictions. In addition, the rapid increase in computational resources, thanks to heavily parallelized computer architectures or clusters of graphic processing units (GPUs), in conjunction with more efficient algorithms for the simulation of biological events has extended the capability of MD simulations to the study of increasingly complex systems, like large protein complexes or membrane embedded proteins. Here, the theoretical foundations of MD and free energy calculation techniques are briefly reviewed. However, as providing a full treatise on all existing techniques goes far beyond the scope of this chapter, only those adopted in my research work are described in some detail. Finally, a brief description of the analyses employed is also provided.

3.1 Statistical mechanics foundations

The microscopic dynamical information about a system (i.e. atomic positions and velocities) can be exploited to study its thermodynamic and kinetic properties. The connection between the macroscopic properties of a system and the spatial distribution and the motion of the atoms composing it is provided by statistical mechanics.[102,103] Within such theoretical framework, an ensemble of microscopic states (also called microstates) underlies, under equilibrium conditions, the thermodynamic state (also called macrostate) of a molecular system, as defined by a set of conserved thermodynamic parameters, such as temperature, T, pressure, P and number of particles, N. Different kinds of statistical mechanical ensembles exist, some of which are listed below.

1) Microcanonical ensemble (NVE): characterized by a fixed number of particles (N), volume (V) and energy (E)
2) Canonical ensemble (NVT): characterized by a fixed number of particles (N), volume (V) and temperature (T)
3) Isobaric-Isothermal ensemble (NPT): characterized by a fixed number of particles (N), pressure (P) and temperature (T)
4) Grand canonical (μVT): characterized by a fixed chemical potential (μ), volume (V) and temperature (T)

From a classical viewpoint, the microstates of a system are simply and fully defined by a vector containing its atomic positions ($\boldsymbol{x}$) and momenta ($\boldsymbol{p}$), which altogether constitute the coordinates of a multidimensional space, referred to as the phase space. For a system of N particles, such space has 6N dimensions and a particular microstate of the system can be also considered as a single point in the phase space. From a purely thermodynamic point of view, all the microstates of an ensemble contribute to the thermodynamics of the system, though with a different weight, as they all correspond to the very same macrostate.

In statistical mechanics, the expectation value of any thermodynamic observable of a molecular system can be expressed in terms of the weighted average, over the entire ensemble of microstates, of a specific microscopic phase space function $f = f(\boldsymbol{x}, \boldsymbol{p})$, as given by the following general equation,

$$\langle L \rangle_{ensemble} = \iint f(\boldsymbol{x}, \boldsymbol{p})\ P(\boldsymbol{x}, \boldsymbol{p})\ d\boldsymbol{x}d\boldsymbol{p}$$

where $f(\boldsymbol{x}, \boldsymbol{p})$ is the microscopic dynamical estimator of the thermodynamic property of interest, L. The weighting function $P(\boldsymbol{x}, \boldsymbol{p})$ represents the ensemble probability distribution function, which generally is a function of the Hamiltonian of the system. For the specific case of the canonical ensemble, the probability distribution function corresponds to the Boltzmann distribution function, that is

$$P(\boldsymbol{x}, \boldsymbol{p}) = \frac{1}{Z} \exp\left(-\frac{\mathcal{H}(\boldsymbol{x}, \boldsymbol{p})}{k_B T}\right)$$

where $\mathcal{H}$ is the Hamiltonian, T is the temperature, k_B is Boltzmann's constant and the normalizing factor $1/Z$ is the reciprocal of the partition function. The partition function is one of the central quantities in equilibrium statistical mechanics and represents a measure of the number of microscopic states in the phase space accessible within a given ensemble. In the specific case of the canonical ensemble, the partition function is an integral, over the entire phase space, of the Boltzmann's factors $\exp(-\mathcal{H}(\boldsymbol{x}, \boldsymbol{p})/k_B T)$ for each microstate i, and has the following form

$$Z = \int \exp\left(-\frac{\mathcal{H}(\boldsymbol{x}, \boldsymbol{p})}{k_B T}\right) d\boldsymbol{x}d\boldsymbol{p}$$

The partition function is an extremely useful quantity in statistical mechanics since it can be used, among other things, to calculate the free energy of a system, through the following relation

$$A = -k_B T \ln(Z)$$

In principle, in order to calculate the ensemble average of a property using eq. (1) one should first obtain the Boltzmann's factors for the entire ensemble of microstates of the system, then calculate the partition function and finally use Z in order to estimate the probability distribution function and thus the expectation value of the property of interest. A convenient method for calculating the

ensemble average is to follow the motion of a single microstate through the phase space as a function of time and then calculate averages over the phase space points sampled along the system's trajectory. The averages over time can be calculated by the following equation

$$\langle L \rangle_{time} = \frac{1}{\tau} \int_{t=0}^{\tau} f(\boldsymbol{x}(t)\, \boldsymbol{p}(t))\, dt \;\approx\; \frac{1}{M} \int_{t=0}^{M} f(\boldsymbol{x}, \boldsymbol{p})$$

where τ is the time, M is the number of time steps of the simulation and $f(\boldsymbol{x}, \boldsymbol{p})$ is the instantaneous value of the microscopic property f.

The statistical mechanics argument which underpins the validity of such time averaging process is the so called ergodic hypothesis. According to such hypothesis, over long periods of time, all the microstates will be sampled. It follows that the time spent by a mechanical system in a region of the phase space of microstates with the same energy is proportional to the volume of that region. When this hypothesis holds true, i.e. for infinitely long trajectories, the probability of visiting the neighbourhood of each point in the phase space converges to a unique limiting value and the ensemble and time averages become independent of the initial conditions and thus equivalent to each other.

3.2 Molecular Mechanics

Molecular dynamics[104] represent one of the election methods to sample the several microscopic configurations of a molecular system characterized by certain thermodynamics conditions, such as temperature, pressure and volume. By following the trajectory of a given microstate, MD simulations allow to follow the time evolution of a system of N particles through the phase space, under the action of their interaction forces. Given a starting phase space point, with initial positions and momenta $\{\boldsymbol{x}(0), \boldsymbol{p}(0)\}$ at the initial instant t_0, the positions and momenta $\{\boldsymbol{x}(t), \boldsymbol{p}(t)\}$ at a following time t for each particle i, are obtained by integrating the Newton's equation of motion for the system:

$$m_i \ddot{x}_i = -\nabla U_i(\boldsymbol{x}), \qquad \forall\, i = 1, \ldots, N$$

where $\boldsymbol{x}$ is a vector representing the configuration of the system, $U(\boldsymbol{x}) : \mathbb{R}^{3N} \rightarrow \mathbb{R}$

is the potential energy function and m_i the mass of the i-th particle. Figure 3.1 shows a general MD simulation algorithm.

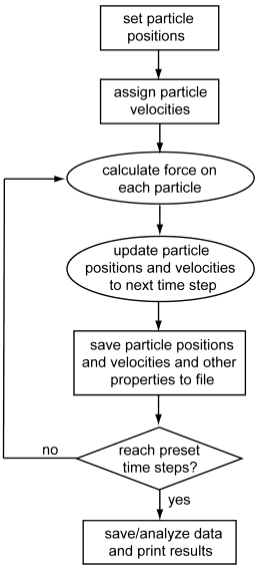


Figure 3.1: Flowchart of a MD simulation.

Actually, a slightly different approach is usually employed in MD simulations, to enforce the influence of a thermostat, i.e. Langevin dynamics. In simple terms, Langevin dynamics differs from usual Newtonian dynamics in the presence of two additional terms to the Newton's equations of motion, aimed at helping to mimic the viscous aspect of a solvent: 1) a term representing the frictional force of a solvent and 2) a term representing a random force. The dynamics of the system can then be written as

$$\begin{cases} d\boldsymbol{x}_t = \mathbf{M}^{-1}\boldsymbol{p}_t\, dt \\ d\boldsymbol{p}_t = -\nabla U(\boldsymbol{x}_t)dt - \gamma\mathbf{M}^{-1}\boldsymbol{p}_t\, dt + \sqrt{2\gamma k_B T}dW_t \end{cases},$$

where $\boldsymbol{x}_t$ and $\boldsymbol{p}_t$ represent the atomic positions and the momenta at time t, $\mathbf{M}$ is the mass matrix, γ is the frictional coefficient, W_t is the Weiner process that underlies the random force (white noise). In explicit solvent simulations, Langevin dynamics with a sufficiently soft damping force and small stochastic forces represents only a mere perturbation of Newtonian dynamics.

Depending on the way the potential energy surface (PES) is computed, MD methods can be mainly divided in two groups: 1) *ab-initio* methods, where the electronic structure is treated explicitly by means of the laws of quantum mechanics (QM), and 2) molecular mechanics (MM) methods (also referred to as classical MD), which, instead, use the laws of classical mechanics to describe the motion of a molecular system. The former methods take explicitly into account the electronic structure, thereby allowing to investigate electronically excited states or chemical bond formation/breaking processes. However, they are normally too computationally expensive for describing large and complex chemical systems. Molecular mechanics instead includes the effect of electrons implicitly and, provided that quantum effects are negligible, can be safely employed to study biological macromolecules in their environment, which would be otherwise hardly treatable via QM methods.

In classical MD, the "mechanical components" of the system, i.e. the atoms, are treated as spheres of a particular radius (typically the van der Waals radius) held together by bonds of a certain strength and experiencing a non-bonding pairwise interaction potential with the other components of the system. The interactions potentials between atoms are empirical, as they are described by suitable combinations of analytical functions, appropriately constructed to fit *ab-initio* calculations, experimental data or a combination of both. The use of such analytical functions (collectively termed force fields) as effective energies allows a fast evaluation of the interatomic interactions, and, as a consequence, larger systems can be sampled in the nanosecond or even microsecond time scale.

It is worth to note that models with a reduced number of interactive centers also exist. Some remarkable instances are the "united atom" representation, where non-polar hydrogen atoms and the heavy atoms they are bonded to are treated as a single particle, or the "coarse grained" representation, where atoms of backbone, side chains or even whole amino acids are condensed into larger "beads". These approaches allow to speed up the computational efficiency and thereby the sampling even further, at the expense of a fully atomistic description of the system .

3.2.1 The Force Field equation

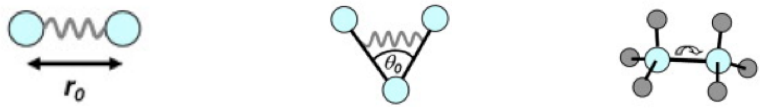
In molecular mechanics, the term force field refers to a set of empirical

analytical functions and corresponding parameters, used to calculate the potential energy surface (PES) of a molecular system as a function of its internal coordinates, $\boldsymbol{x}$. All the components of the force field describe a particular pairwise interatomic potential. The parameters for each single interaction term in the force field are conveniently derived to be, as much as possible, specific for a given atomic type, i.e. a practical concept adopted to represent atoms with same hybridisation or in similar chemical contexts. Currently there are many different force fields available for studying molecular systems. Among those parametrized and optimized to be accurate and transferable for proteins and nucleic acids, AMBER[105] and CHARMM[106] are two of the most important. They differ mostly in the parametrization approaches. AMBER is parametrized according to structural and thermodynamic data mostly coming from quantum mechanical calculations, while CHARMM uses prevalently experimental data.

$$U(\boldsymbol{x}) = U(\boldsymbol{x})_{bonded} + U(\boldsymbol{x})_{non\ bonded}$$

$$U(\boldsymbol{x})_{bonded} = U(\boldsymbol{x})_{bonds} + U(\boldsymbol{x})_{angles}\, U(\boldsymbol{x})_{dihedrals} =$$

$$= \sum_{bonds} \frac{1}{2} k_b (r - r_{eq})^2 + \sum_{angles} \frac{1}{2} k_a (\theta - \theta_{eq})^2 + \sum_{dihed} k_\phi [1 + \cos(n\phi - \delta)]$$



$$U(\boldsymbol{x})_{non\ bonded} = U(\boldsymbol{x})_{Coulomb} + U(\boldsymbol{x})_{van\ der\ Waals} =$$

$$= \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{i,j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]$$

Figure 3.1: Functional form of a general force field.

The single interatomic potentials can be conveniently divided into two groups, representing the bonded and non-bonded interactions. The bonded terms are associated with atoms connected by one up to three covalent bonds, while the non-bonded terms account for the interactions with all other non-bonded atoms (Fig. 3.1). Additionally, specific terms describing the hydrogen bonds or the correlations among different degrees of freedom (DOF) can be present. For instance, a harmonic potential term is often used to treat out-of-plane distortions occurring within aromatic rings.

Concerning the bonded terms, the bond-stretching and angle-bending terms are usually modelled harmonically, keeping the bond lengths and angles near their equilibrium values (r_{eq}, θ_{eq}) with the associated force constants k_b and k_a, respectively. The dihedral terms represents the rotation that occur about a bond, leading to the changes in the position of atoms i and $i + 3$. Since these terms are periodic in nature, a sinusoidal function is used to model them. The parameters k_ϕ, n and δ are the force constant, periodicity and phase, respectively.

The non-bonded interactions in Eq. 3 describe both inter-molecular and intra-molecular interactions. The first term describes the Coulombic or electrostatic interaction between atoms i and j endowed with partial atomic charges q_i and q_j, and separated by the distance r_{ij},, with dielectric constant ϵ_0. The term in square bracket is used to treat the van der Waals (vdW) interaction and is referred to as the Lennard-Jones (LJ) 6-12 potential. Such potential is made up by two terms. The $1/r^6$ term represents the London's dispersion interactions (i.e. instantaneous dipole-induced dipole) while the $1/r^{12}$ term is an arbitrary potential adopted to represent the repulsion force between two atoms when their electron clouds overlap. The well depth, ϵ_{ij}, indicates the magnitude of the London's dispersion interaction between atoms i and j and σ_{ij} is the threshold distance below which the LJ interaction energy goes towards plus infinity.

Once the 3D structure and all the required parameters are known, the energy and the gradient of the system can be calculated from the combination of the analytical functions reported above.

3.3.2 The choice of the time step

For many-body systems such as molecular systems, the integration of

Newton's equations of motion cannot be carried out analytically. Instead, it must be performed numerically, for instance through the finite difference method. Ultimately, the numerical solution of a set of second order differential equations is the core engine of any molecular dynamics package. The idea at the basis of the method is to replace derivatives with incremental ratios, by breaking down the integration of Newton's equations in many fixed stages, each separated in time by a small time step δt.

First initial position and velocities have to be assigned. The initial configuration of the system is usually obtained via experimental or theoretical models, or a combination of both. The initial velocities are usually generated through a random number generator, according to the Maxwell-Boltzmann distribution at a certain temperature *T*.

$$T = \frac{1}{3N}\sum_{i=1}^{N}\frac{|p_i|}{2m_i} \qquad p(v_{xi}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{m_i\, v_{xi}^2}{k_B T}\right]$$

Once starting positions and velocities have been assigned, the total force on each particle in the configuration at the time t is calculated as the vector sum of its interaction with other particles. From the forces it is possible to determine the accelerations of the particles, which are then combined with the positions and velocities at the time t to calculate the positions and velocities at the time $t + \delta t$. The force is assumed to be constant during the time step (hence no third derivative of the position is considered). The forces on the particles in their new positions are then determined, leading to new positions and velocities at the time $t + 2\delta t$, and so on. Many algorithms are available for integrating the equations of motion, based on the assumption that positions, velocities and accelerations can be approximated as Taylor series expansions, generally truncated to the second order. The main difference among the algorithms lies in the way they calculate the aforementioned quantities, and particularly whether or not they calculate all of them for the same time step (like the *leap-frog* algorithm) and whether they calculate them more than once for each time step (like in the *velocity Verlet* algorithm, with the accelerations, or like the *Verlet* algorithm, with the positions). Each method has its advantages and drawbacks in terms of precision and storage requirements (for example the leap-frog algorithm has the obvious disadvantage that positions and velocities are not synchronized; this means that it not possible

to calculate the kinetic energy contribution to the total energy at the same time as the positions are defined, from which the potential energy is determined).

Whatever the algorithm one use to integrate the equations of motion, the step that mostly affects the efficiency of any MD algorithm is definitely the calculation of the interactions between each particle. As this step represents the computational bottleneck of the algorithm, it can severely limit both the sampling duration (and so the data collection) and the dimension of the system that can be investigated. In this respect, the choice of the time step is pivotal. Indeed, it is very important to find the right balance between accuracy, by using, as much as possible, a small time step in order to preserve energy and momenta and avoid errors and artifacts in the simulation, and phase space exploration, which instead would benefit from the use of a larger time step. The maximum time step that can be used safely is dictated by the frequency of the fastest motion in the system, as it should be approximately one-tenth that time. For molecular systems, these are the stretching vibrations involving the hydrogen atoms, (approximately 10 femtosecond) leading to time steps of 1 femtosecond (fs). One the one hand, such a small time step ensures good accuracy for the sampling of the statistical ensemble; on the other hand, this requirement is a severe restriction, as it compels a huge number of steps to sample the phase space efficiently. However, the time scale of these high frequency motions is orders of magnitude smaller than that of larger and more functionally relevant motions, such as conformational changes or folding events, which occur on time scales orders of magnitude much larger. Therefore, since these motions are usually of relatively little interest and have a minimal effect on the overall behavior of system like proteins or nucleic acids, they are commonly frozen out applying some *constraints* with respect to the heavy atoms they are bonded to. Such strategy allows to increase the time step (for example to 2 fs), thus easing the exploration of the phase space.

Moreover, in order to add even more flexibility, some molecular dynamics codes allow the use of the so called “multiple time step”. Through this method it is possible to calculate different sets of non-bonded interactions with different frequencies, for example with a smaller time step for interactions involving atoms inside a cutoff distance and a longer time step for the ones outside the cutoff distance. This is based on the fact that certain interactions evolve more rapidly with time than other ones.

3.2.3 Periodic boundary conditions

As the size of a typical simulated system is limited, a practical solution must be found to deal with artificial boundary effects. There are two common ways for treating the boundaries: *periodic boundary conditions* and *spherical boundary conditions with continuum based models*. Each of these methods has its advantages and drawbacks and it is more suited than the other to the study a particular problem. Generally, for the simulation of biological macromolecules, the first one is the most widely adopted. In periodic boundary conditions (PBC), the simulation cell is periodically replicated throughout the three dimensions of the space to form an infinite lattice of images, or copies (Fig. 3.2). By, this way, any atom close to the boundaries of the simulation cell can interact with another atom in the neighbouring cell, provided it is within a cut-off radius, r_c. To avert ambiguities, i.e. the possibility to interact with two images of the same atom, r_c cannot be greater than half the width of the cell, so that an atom can interact with only one image of another atom. As a result, the interactions are always calculated only with the closest image range. This is known as the minimum image convention. In case, during the course of the simulation, an atom (or a molecule) should leave the central box, one of its periodic images will enter through the opposite face, to keep the number of atoms within the simulation cell constant. In order to avoid spurious correlation of motions between adjacent images, the unit cell size must be chosen to be sufficiently large when using PBC.

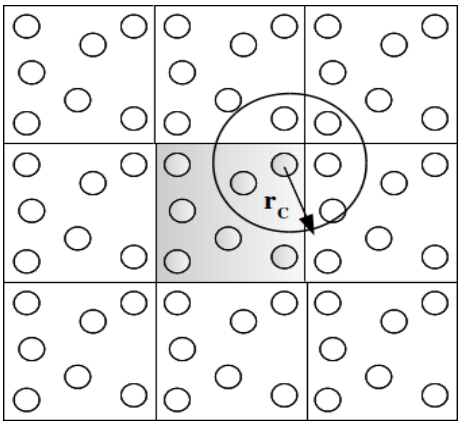


Figure 3.2: Periodic boundary conditions in two dimensions. The central simulation cell is replicated infinitely in both dimensions.

3.2.4 Long range interactions

As mentioned above, the most computationally expensive part of MD simulation is the evaluation of the non-bonded interaction terms, since these represent the most numerous interactions to calculate. Indeed, the direct assessment of the non-bonded interactions between all atom pairs has the complexity of $O(N^2)$, where N is the number of atoms. However, it is commonly accepted that, in order to avoid the expensive energy calculation between all atom pairs, it is possible to safely define a cut-off distance (r_c), smaller than half of the simulation box, beyond which long range interactions can be evaluated in an approximated fashion, without losing appreciable accuracy. Such approximations separately cope with Lennard-Jones (LJ) interactions and long-range electrostatics interactions.

In the case of the LJ interactions, the interactions beyond the cut-off radius are simply ignored. The validity of this approach is granted by the fact that LJ have an extremely strong dependence on the distance between the interacting centres ($1/r^6$ for the attractive term), so they can be safely neglected beyond a certain threshold distance, r_c. However, since this may lead to discontinuities at the cut-off distance, the non-bonded interaction term can be also multiplied by or added to appropriate shifting or switching functions, respectively, to ensure it goes smoothly to zero at r_c. Moreover, in order to further reduce the computational cost of finding all atom pairs within the cut-off distance for each step, it is possible to store all those atoms contained within a sphere of radius $r_b > r_c$ from a given atom within lists, which can then be updated at a lower rate throughout the simulation. Such method of book keeping the LJ interactions is called the neighbour lists method (Figure 3.3).

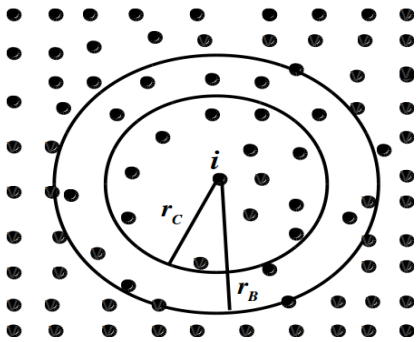


Figure 3.3: Neighbour list creation during simulation.

At variance with the LJ interactions, the long-range electrostatic interactions display a less marked dependence on the distance between atoms ($1/r$) and, therefore, their neglect introduces severe approximations, especially for highly charged systems. Hence, alternative techniques have been developed and now are widely used, such as the particle mesh Ewald (PME) method.[107] The basic principle of this method is to replace the direct summation of the pairwise long-range electrostatic interactions with a summation in the Fourier (reciprocal) space. The PME method makes an implicit assumption of periodicity and therefore it requires the imposition of spatial periodic symmetry to system in study. However, as it will be explained in the next paragraph, this is the standard situation for the simulation of biomolecules.

Under PBC, the electrostatic energy associated with the central cell due to the Coulombic interaction of atoms located within the central cell and atoms in the replica cell is defined as

$$U_{elec} = \frac{1}{4\pi\varepsilon}\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n\,\epsilon\,Z^3}\frac{q_i q_j}{|r_{ij}nL|}$$

The Ewald summation is based on splitting the interaction $1/r$ into two series

$$\frac{1}{r} = \frac{g(r)}{r} - \frac{1-g(r)}{r}$$

where $g(r) = \mathrm{erfc}(\alpha, r)$ and α is called the Ewald splitting parameter. This results in the well known Ewald formula for the energy of the main cell

$$U = U^{(r)} + U^{(K)} + U^{(self)} + U^{(dipolar)}$$

where $U^{(r)}$ represents the real space contribution, $U^{(K)}$ the reciprocal space contribution, $U^{(self)}$ the self interaction energy and $U^{(dipolar)}$ the dipolar correction. In practice the sums for $U^{(r)}$ and $U^{(K)}$ are evaluated defining cut-offs given by r_{cut} and k_{cut}.

The summation in the reciprocal space is achieved upon converting a system

of particles (in this case atoms) into a charge density field, represented on a grid (or "mesh") and then evaluate the potential for the grid points using the efficient Fast Fourier Transform (FFT) method, which substantially reduces the time needed to compute the reciprocal contribution $U^{(K)}$.

The condensing of the charges into density field points makes the PME method more efficient for systems with "smooth" variations in the density. However other methods to treat systems where the density is characterized by larger fluctuations, like the fast multipole method, also exist.

The schemes presented above have revolutionized the field of biomolecular simulation in the past decade, since their computational complexity is only $O(N \log(N))$, a dramatic computational saving with respect to the direct value of $O(N^2)$.

3.3 Enhanced Sampling and Free Energy Calculations

In order to get a satisfactorily accurate estimate of the properties of a system as well as the free energy changes associated with its dynamics, the phase space must be appropriately explored. Indeed, as mentioned above, molecular dynamics is based on the assumption that the ergodic hypothesis is verified. However, there exist many systems that behave as non ergodic in the time scale of standard MD simulations, a phenomenon referred to as quasi-nonergodicity.[108] When a simulation cannot be carried out over sufficiently long time periods to compensate for quasi-nonergodicity problems, the statistical sampling issuing from the phase space exploration could be incomplete and consequently the ergodic assumption could not hold. Under such circumstances the time averages of the microscopic properties cannot be replaced by averages over the ensemble. Moreover, in these instances, statistical averages might exhibit strong dependence from the initial conditions, i.e. different MD simulations of the same system could not yield the same results.

To be specific, a poor description of the properties and the dynamics of a system will ensue whenever the system itself cannot visit all its energetically relevant configurations. For instance, in the case of a free energy calculation, this may lead to large errors in the estimate of the relative free energy between two configurations of a system, even up to impede a correct discrimination between metastable and transition states. As will be outlined in the following paragraph,

the situation is particularly problematic in the case of large and complex molecular systems, such as biomolecules.

3.3.1 The problem of rare events

The main reason behind frequent quasi-nonergodicity problems with biomolecular systems lies in the fact that such systems are characterized by a very rugged free energy landscape.[109] As most of the energetically relevant configurations are, in practice, often separated by relatively high free energy barriers, a system can remain kinetically trapped in one or few free-energy basins under normal conditions of Boltzmann sampling. This can result in a very slow diffusion through other phase space important regions.

Many complex biophysical processes, including folding/unfolding, extended conformational changes and ligand binding events, are represented by transitions between phase space volumes and, as such, their modeling features a very high chance to run into quasi-nonergodicity problems. Indeed these processes, which are often referred to as “rare events”, usually occur in the range of microseconds to seconds, while standard MD simulations, unfortunately, can be extended only up to hundreds of nanoseconds. Furthermore, even in case the important phase space volumes are all connected by low free-energy regions but these regions, termed as “entropy bottlenecks”, are very narrow and hence rarely sampled, problems of quasi-nonergodicity may persist as well. This eventually results in a phase space volume covering largely insufficient for a reliable characterization of the problem under study. Two different routes can be pursued to deal with the problem of time scale. The first one consists in running extremely long MD simulations in order to collect enough statistics of the process under investigation.[110-112] However, especially in the case of larger systems like protein complexes or membrane proteins, it is frequently impractical to follow such a “brute force” approach. The second one is based on the employment of the so-called “enhanced sampling” methods. During the last years, several enhanced sampling methods have been developed to successfully cope with the rare-event problem and achieve a more thorough exploration of all the energetically important phase space regions.[113] In the next paragraphs, an overview of the two main strategies which constitute the basis for those methods that nowadays are routinely applied to the study of complex biomolecular processes is given.

Then, the specific enhanced sampling method employed in this thesis work is presented, preceded by a brief introduction to the thermodynamic integration formalism for calculating the free energy, from which the method naturally arises.

3.3.2 Escaping from Boltzmann sampling

Among the many so-called variance reduction statistical techniques, two strategies have proved to be particularly effective in dealing with quasi-nonergodicity and, as such, employed in enhanced sampling methods, namely stratification[114] and importance sampling.[115] In fact, almost all methods used to improve sampling efficiency, especially for free energy calculations purposes, rely on one of these strategies, or on their thoughtful and creative combination.

Stratification, sometimes also called multistage sampling, is a strategy for distributing samples so that all the relevant parts of a population required to obtain an accurate probability distribution function, $P(\boldsymbol{x})$, are adequately sampled. In an unstratified sampling method, all the samples are generated from the same probability distribution function, which might vary greatly in Ω, the domain of accessible microstates for the energy levels compatible with a given macrostate. Conversely, in a stratified method the microstates population is first divided into homogeneous subgroups, called strata or windows, and then the so-obtained subgroups are sampled independently. Specifically, the domain Ω is first partitioned into a number of disjoint regions Ω_i, such that their union covers the whole domain Ω. In the region Ω_i, $\boldsymbol{x}_i$ is sampled according to $P_i(\boldsymbol{x}_i)$, equal to $P(\boldsymbol{x})$ in this region. This process displays the obvious advantage that every strata is sampled, even if it is associated with a very low $P(\boldsymbol{x})$, conversely to what should happen in an unstratified sampling method, wherein it would be unlikely visited. The end result of stratification is the improvement of the representativeness of the sample and a reduction of the sampling error, leading to a lowered variance on the estimate of any function $f(\boldsymbol{x})$ averaged over Ω with the probability measure $P(\boldsymbol{x})$. To illustrate how stratification works in the context of free energy calculations, let us consider the transformation of a system from a state 0 to a state 1 described by a parameter λ. The free energy profile for the transformation can be written as:

$$A(\lambda) = -\beta^{-1} \ln P(\lambda) + A_0$$

where $A(\lambda)$ is the free energy at state λ, $P(\lambda)$ is the probability distribution to find the system at state λ along the transformation coordinate, β is the thermodynamic beta, $\beta = 1/k_B T$, and A_0 is a constant. $P(\lambda)$ can be defined as

$$P(\lambda) = \frac{\int \exp(-\beta\, \mathcal{H})\, \delta(\lambda - \lambda(\boldsymbol{x}))\, \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{p_x}}{\int \exp(-\beta\, \mathcal{H})\, \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{p_x}}$$

Let us further assume that these two states are separated by a high energy barrier that corresponds to a value of λ between λ_0 and λ_1. Transition between 0 and 1 would thus be rare as the intermediate regions would be unlikely sampled. Therefore, the free energy estimated from unstratified computer simulations would converge very slowly to its limiting value, irrespective of the initial conditions. If, however, the full range of λ is partitioned into a number of small intervals, and each of these intervals is sampled independently, it is possible to recover the complete $P(\lambda)$ and reliably estimate free energy differences, with great savings of computer time.

Importance sampling is another, highly successful variance-reduction strategy. The idea behind it is that certain regions in phase space are important for estimating the quantity of interest (e.g. from the energetic point of view), even though these regions might have low probability of being visited. It is thus advantageous to choose a sampling distribution from which these "important" regions are sampled more frequently than they would be from the true distribution, thereby improving sampling uniformity. From a MD perspective, this means helping the system to leave from a free energy basin in which it is trapped by applying a sort of "bias" to the system. In general, the correct bias, which usually cannot be known a priori as it strongly depends on the system and the problem under study, needs to be guessed beforehand or can be gradually improved using an iterative refinement process. The latter procedure is that commonly adopted and consists in first running a short simulation to get reasonable estimate of the bias and then applying the so-estimated bias to the system. With the first bias, the sampling is improved and a more accurate approximation of the bias can be obtained. The bias can be then gradually improved in this fashion. However, if this approach was applied directly in a simulation, it would yield a biased estimator, $P'(\lambda)$.

$$P'(\lambda) = \frac{\int \eta(\lambda) \exp(-\beta \mathcal{H}) \, \delta(\lambda - \lambda(\boldsymbol{x})) \, \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{p}_{\boldsymbol{x}}}{\int \eta(\lambda) \exp(-\beta \mathcal{H}) \, \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{p}_{\boldsymbol{x}}}$$

Because $P'(\lambda)$ differs from the correct Boltzmann distribution for the system, the methods based on sampling $P'(\lambda)$ are often called non-Boltzmann sampling techniques. The results of the simulation obtained using the modified distribution can, however, be properly corrected (i.e. reweighted) to ensure that the estimator is unbiased and so the true Boltzmann distribution recovered. The weight is given by the likelihood ratio of the true distribution to the biased simulation distribution. The basic idea of importance sampling can be illustrated easily in the example, described above, of the transformation from 0 to 1 along λ. In lieu of sampling from the true probability distribution, $P(\lambda)$, it is possible to design simulations in which λ is sampled according to $P'(\lambda)$. Obviously, the latter should be chosen so that it is more uniform than $P(\lambda)$. The relation between the two probabilities may then be expressed as follows:

$$P'(\lambda) = P(\lambda) \exp[\beta\eta(\lambda)]$$

where $\eta(\lambda)$ is a weighting factor that depends on the value of λ. Next, the difference in free energy, ΔA, can be expressed in terms of $P'(\lambda_1)$ and $P'(\lambda_0)$ derived from the biased simulation:

$$\Delta A = -\beta^{-1} \ln \frac{P(\lambda_1)}{P(\lambda_0)} = -\beta^{-1} \ln \frac{P'(\lambda_1)}{P'(\lambda_0)} + \eta(\lambda_1) - \eta(\lambda_0)$$

The fundamental issue in implementing importance sampling in simulation is the choice of the biased distribution, or, equivalently, the weighting factor, η. A variety of ingenious techniques that lead to great improvement in the efficiency and accuracy of free energy calculation have been developed for this purpose.

It is now well-established that the ergodic hypothesis typically tends to break down locally rather than globally, along few particular degrees of freedom (DOF) characterized by a slow dynamics, for which barrier-crossing is disfavored. Therefore, many enhanced sampling methods are based on improving the sampling along a selected number of DOFs of the system, often referred as

collective variables. These collective variables are frequently generalized coordinates, i.e. functions of the system's atomic coordinates. Thus, if ξ is a set of d functions of the atomic coordinates $\boldsymbol{x}$ of the system, the additional time-dependent bias, is applied exclusively along the collective variables ξ and the Newton's equations of motions are modified accordingly. For instance, if the bias has the form of a biasing potential $V(\xi(\boldsymbol{x}), t)$ the equations of motion can be rewritten as:

$$M\ddot{R} = -\frac{\partial U(\boldsymbol{x})}{\partial \boldsymbol{x}} - \frac{\partial V(\xi(\boldsymbol{x}), t)}{\partial \boldsymbol{x}}$$

Once all the regions of the collective variable space have been explored satisfactorily, it is possible, if the sampling has been carried out reversibly, to recover the free energy profile along ξ, by means of the equation above. If the selected collective variable space represents a good descriptor of the transformation in study, it is possible to identify the energetically important configurations of the system associated to the transformation itself. This allows to discriminate between actual free energy minima and metastable states, as well as to identify transition states, e.g. in order to obtain an estimate for the kinetic rate of the process.

3.3.3 The Thermodynamic Integration formalism

The free energy along a collective variable can be also seen as a potential resulting from the average force, that is the negative of the gradient of this potential, acting along the collective variable itself. Such potential is frequently referred to as potential of mean force (PMF). The PMF was historically defined as the reversible work required to bring two solvated particles from infinity to a contact distance.[102] However, later this has been generalized to a plethora of collective variables that go far beyond the simple distance between atoms or molecules.[116] This way of seeing the free energy as a potential deriving from an average force is employed by a class of widely used approaches to calculate free energy profiles along a chosen ξ, based on the thermodynamic integration (TI) formalism.[108,117-121] In this framework, the "thermodynamic" force, i.e.

the first derivative of the free energy with respect to one or more collective variables, is calculated directly, at a fixed ξ values, and subsequently integrated to recover the free energy profile (i.e. the potential) along ξ. Specifically, the thermodynamic force along a generalized collective coordinate ξ can be expressed as:

$$\frac{\mathrm{d}A}{\mathrm{d}\xi} = \frac{\int \frac{\mathrm{d}\mathcal{H}}{\mathrm{d}\xi} \exp(-\beta\,\mathcal{H})\,\mathrm{d}q_1 \ldots \mathrm{d}q_{N-1}\mathrm{d}p^{\xi} \ldots \mathrm{d}p^{q}_{N-1}}{\int \exp(-\beta\mathcal{H})\,\mathrm{d}q_1 \ldots \mathrm{d}q_{N-1}\mathrm{d}p^{\xi} \ldots \mathrm{d}p^{q}_{N-1}} = \left\langle \frac{\mathrm{d}\mathcal{H}}{\mathrm{d}\xi} \right\rangle_{\xi} = -\langle F_{\xi} \rangle_{\xi}$$

where $q_1 \ldots q_{N-1}$ are the coordinates arising from the coordinate transformation process from Cartesian to generalized coordinates and $p^q_1 \ldots p^q_{N-1}$ are their conjugated momenta. Such an average over the phase space isosurface characterized by a specific ξ value corresponds naturally to the "generalized" average force acting on ξ itself.

Note that while in the case of a general collective variable ξ the mean force is a well-defined quantity, its expression as an ensemble average of an instantaneous force F_ξ exerted on ξ is not unique. In simulations, the choice of a convenient expression for $\langle F_\xi \rangle_\xi$ is driven by several practical considerations, as ease of implementation and numerical behavior, such as variance. A valid expression for the average force[119] is

$$\left\langle \frac{\mathrm{d}\mathcal{H}}{\mathrm{d}\xi} \right\rangle_{\xi} = \left\langle \nabla U \cdot \frac{\partial \boldsymbol{x}}{\partial \xi} - k_B T \frac{\partial \ln|J|}{\partial \xi} \right\rangle_{\xi} = -\langle F_{\xi} \rangle_{\xi}$$

where $|J|$ is the determinant of the Jacobian matrix $(\partial_i \Xi_j^{-1})_{(i,j)}$ upon the transformation from Cartesian to generalized coordinates. That is, $\boldsymbol{\Xi}: \mathbb{R}^{3N} \rightarrow \mathbb{R}^{3N}$, with $\Xi_1 = \xi$ and components Ξ_i for $i > 1$ are generalized coordinates of no particular physical significance, but are necessary to the mathematical framework. The Jacobian $|J|$ measures the change in volume element between $\mathrm{d}\xi \mathrm{d}\boldsymbol{q} \mathrm{d}p^{\xi} \mathrm{d}\boldsymbol{p}^{q}$. The derivative of A is therefore the sum of two contributions: the mechanical forces acting along ξ, $(\nabla U \cdot \partial \boldsymbol{x}/\partial \xi)$, and the change of volume element along ξ, $(-k_B T\, \partial \ln|J|/\partial \xi)$. The latter is effectively an entropic contribution. The direction along which the forces are projected is represented by the vector $\partial \boldsymbol{x}/\partial \xi$, which is

called the "inverse gradient".[122] As the gradient of ξ can be seen as the changes in ξ corresponding to infinitesimal changes in $\boldsymbol{x}$, the inverse gradient is the vector along which a change in ξ is translated to the atomic Cartesian coordinates, other generalized coordinates being constant ($\boldsymbol{\Xi}_{i,j>1}$). Different choices of $\boldsymbol{\Xi}$ correspond to different inverse gradients and hence to different ways of projecting the Cartesian forces, $-\nabla U$, onto the collective variable.

Unfortunately, in several cases, the numerical computation of the partial derivative of the Jacobian $|J|$ with respect to ξ may be very difficult, as it requires the second derivative of $\boldsymbol{\Xi}^{-1}$, whose analytical derivation may be cumbersome. Moreover, obtaining an explicit coordinate transformation for a full set of generalized coordinates (which include the transition coordinate ξ) can be exceedingly difficult, especially if ξ is a function of the coordinates of several particles. It is therefore desirable to have an expression in which the explicit definition of generalized coordinates other than ξ is not required. Such expression is given by

$$\frac{\mathrm{d}A}{\mathrm{d}\xi} = \left\langle \nabla U \cdot \frac{\mathbf{w}}{\mathbf{w} \cdot \nabla \xi} - k_B T \, \nabla \cdot \frac{\mathbf{w}}{\mathbf{w} \cdot \nabla \xi} \right\rangle_{\xi}$$

This equation has been derived by den Otter[123] and Ciccotti et al.[124] and is general, in the sense that it holds for an arbitrary vector field **w,** provided that $\mathbf{w} \cdot \nabla \xi \neq 0$. There is a direct connection between the choice of the generalized coordinates and **w**, i.e. the choice of **w** is arbitrary just as the choice of $q_1 \ldots q_{N-1}$. For instance, by choosing $\mathbf{w} = \partial \boldsymbol{x} / \partial \xi$, the equation above is recovered. Instead, if $\mathbf{w} = \nabla \xi = \partial \xi / \partial \boldsymbol{x}$, then the expression for the derivative of the free energy becomes

$$\frac{\mathrm{d}A}{\mathrm{d}\xi} = \left\langle \frac{\nabla U \cdot \nabla \xi}{|\nabla \xi|^2} - k_B T \, \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \right\rangle_{\xi}$$

which depends solely on the collective variable ξ.[125] Thus, through the equation above, changes in ξ may be propagated along an arbitrary direction in Cartesian coordinates, without explicitly define a complete set of generalized coordinate. This idea can also be extended to a multidimensional collective variable case, provided that certain requirements are satisfied.[124]

Actually, constraining the system at fixed values of the collective variable is not a prerequisite to calculate the thermodynamic force but, in a few instances, may even be detrimental, as it can reduce sampling efficiency for fixed ξ values. Subsequently it was showed that unconstrained TI formalism could be efficiently combined with stratification and importance sampling strategies into a simulation scheme that yielded uniform sampling along the order collective variable. This approach was called the adaptive biasing force (ABF) method.

3.3.4 The Adaptive Biasing Force method

The ABF method[122,126-129] is a numerical scheme based on stratification and importance sampling, aimed at calculating PMF profiles along a collective variable of interest. The ABF method relies on computing the average force acting on a system and then using the aforementioned relation between the free energy and the average force to yield the PMF profile. The average force on ξ is estimated as a mean over the instantaneous forces on ξ, which, in turn, can be decomposed into the sum of the average force itself (which depends only on the value of ξ) and a random force with zero average, reflecting the fluctuation of all other degrees of freedom.

The pulsing core of ABF relies upon the simple idea of adding to the equations of motion of the system, a continuously updated biasing force, with equal magnitude and opposite sign of the average force acting along ξ. This cancels out the estimate of the average force, erasing the ruggedness of the free energy landscape and thus helping the system to escape from kinetic traps and to sample uniformly all the values of the collective variable. Then, as a simulation progresses, the algorithm adaptively refines the estimate for this average force as new configurations of the system for a given value of ξ are explored. That yields, in the long time limit, an Hamiltonian devoid of an average force acting along the collective variable. Once convergence of the biasing force to the average force at equilibrium is achieved, the dynamics of ξ corresponds to a random walk with zero mean force, as only the fluctuating part of the instantaneous force, i.e. the random force, remains. Thereby the sampling proceeds uniformly on a flat free energy surface, driven exclusively by the diffusion properties of the system.

In ABF generic one dimensional implementation, the collective variable, ξ, connecting two end points, is partitioned into M windows, each one divided into k

equally sized bins of width $\delta\xi$. In the course of the simulation, every time ξ falls within a certain bin, the force samples for that particular bin are accrued in the form of a running time average. Hence, the approximation to the average force in bin k after N_{step} molecular dynamics steps is just the simple, unweighted average of all force samples in that bin

$$\overline{F}_{\xi}(N_{step}, k) = \frac{1}{N_{step}^{k}} \sum_{\mu=1}^{N_{step}^{k}} F_{\mu}^{k}$$

provided that the bin has already been visited at least once. N_{step}^{k} is the number of samples accrued in bin k after N_{step} steps and F_{μ}^{k} abbreviates the μ-th force sample in this bin, which typically obeys a normal distribution. This approach would work well for large N_{step}^{k}.

When only a few samples are available in a given bin k, the running average may be a poor estimate of the actual average force in this bin. Similarly, adding additional samples might markedly change $\overline{F}_{\xi}\left(N_{step}, k\right)$. In either case, large variations in the running estimate of the average force are undesirable, as they may lead to nonequilibrium effects and systematic bias of the calculation. This would slow down the convergence of the algorithm and reduce the efficiency of the method. In order to control potential initial nonequilibrium effects, a ramp function can be added, which reduces the variations of the external force applied in a given bin from one step to the next. Thus, the external force applied to the system can be chosen to be equal to:

$$-R\left(n^{k}(N)\right)F_{\xi}^{k}(N)$$

$$R(n) = \begin{cases} n/N_0 & \text{if } n \leq N_0 \\ 1 & \text{if } n > N_0 \end{cases}$$

Through the ramp function, the biasing force at time t is applied in full only if the number of samples N_t^k is above a threshold, N_0, which should be chosen relatively small. In general, it is important to have a good balance between rapidly improving the sampling of the system, which requires a small N_0, and avoiding large nonequilibrium effects, which requires a large N_0.

As mentioned above, the method is based on the discretization of the collective variable, ξ, in bins of width $\delta\xi$ in which statistical information is accrued in the course of the simulation. It has to be noted that the choice of the bin width is also of importance. Indeed, on the one hand, too small values of $\delta\xi$ require longer simulation times to collect sufficient force statistics in every bin. On the other hand, too large values of $\delta\xi$ prevent capturing variations of the average force on short length scales, which may have adverse effect on the accuracy of the subsequent integration operation.

For a sufficiently large N_{step}, $\overline{F}_{\xi}(N_{step}, k)$ approaches the correct average force in each bin. Then, the free energy difference, ΔA_{ξ}, between the end point states (and hence the PMF profile) can be estimated simply by way of summing the force estimates in individual bins

$$\Delta A_{\xi} = -\sum_{i=1}^{k} \overline{F}_{\xi}(N_{step}, k)\, \delta\xi$$

The peculiarity of ABF as compared to adaptive potential enhanced sampling schemes which are based on probability distribution functions, is due to the fact that forces, at variance with probabilities, are local properties and therefore they can be readily estimated without the need to sample broad ranges of the order parameter. In simpler terms, the method features the absence of any requirements for prior knowledge about the system's free energy landscape, in order to estimate a correct bias, with potential benefits in terms of the efficiency of the algorithm.

3.4 Trajectory Analyses

3.4.1 RMSD analysis

The root mean square deviation (RMSD) is a measure of the average distance between the atoms of a configuration at a given time and a reference structure obtained after an optimal rigid body superposition. The RMSD is calculated as follows,

$$\text{RMSD}\left(V_t, W_{ref}\right) = \sqrt{\frac{1}{N}\sum_{1}^{N}(\boldsymbol{x}_{i,t} - \boldsymbol{x}_{i,ref})^2}$$

where $\boldsymbol{x}_{ref}$ is the reference structure, $\boldsymbol{x}_t$ is the structure at a given time, N is the number of atoms, while $\boldsymbol{x}_{i,t}$ and $\boldsymbol{x}_{i,ref}$ are the coordinates of the i -th atom of structure $\boldsymbol{x}_t$ at time t and $\boldsymbol{x}_{ref}$. RMSD value is expressed in units of length, typically Angstroms (Å). In MD trajectory analysis, RMSD is used as a measure of the stability of the molecule, upon equilibration. In a typical plot of RMSD vs time, one can see that the value increases initially due to the relaxation of the molecule from the starting model and reaches a plateau after a few nanoseconds. This indicates that the molecule, once left the initial state, fluctuates around a defined configuration close to a local free energy minimum. It is this stable part of the trajectory that is usually considered for the analysis of the structural property of the molecule. The RMSD analysis presented in the results have been calculated for the backbone atoms.

3.4.2 RMSF and b-factors analysis

The root mean square fluctuation (RMSF) is a measure of the amplitude of fluctuation in Angstroms with respect to the time averaged position and is used to assess the flexibility of different regions of a molecule. The RMSF is calculated as

$$\text{RMSF} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\boldsymbol{x}_t - \langle \boldsymbol{x} \rangle)^2}$$

where $\boldsymbol{x}_t$ is the position of an atom at time t, N is the total simulation time, while $\langle x \rangle$ is the average position of the atom. It is also possible to calculate the RMSF for groups of atoms, e.g. for an entire residue, by simply averaging the RMSF values of the atoms composing a particular group. Moreover, in a few instances it may be more convenient to calculate a parameter measuring the atomic fluctuations that is more similar to the crystallographic observable for thermal

fluctuations, i.e. the b-factors, by means of the following equation

$$b - factor = \frac{8}{3}\pi^2(\mathrm{RMSF})^2$$

3.4.3 Secondary structure analysis

Protein secondary structure has been evaluated following the DSSP method by Kabsch & Sander,[130] as implemented in the ptraj program of Amber package.[131] This method takes atomic coordinates as input and assigns eight different secondary structure elements using hydrogen bonds alone. DSSP defines a hydrogen bond where the bond energy is below -0.5 kcal/mol based on a Coulomb approximation of the hydrogen bond energy. The output data from secondary structure analysis is usually depicted in a graphical representation with different colors for each secondary structure element.

3.4.4 Dynamical Cross Correlation analysis

The Dynamical Cross Correlation (DCC) analysis allows to assess the extent to which the atomic displacements of a system are correlated with each other by examining the magnitude of the pairwise dynamical cross-correlation coefficients, which are computed as:

$$\mathrm{DCC}_{a,b} = \frac{\langle \Delta r_a(t) \Delta r_b(t) \rangle_t}{\sqrt{\langle \|\Delta r_a(t)\|^2 \rangle_t} \sqrt{\langle \|\Delta r_b(t)\|^2 \rangle_t}}$$

where Δr_a represents the displacement from the mean position of the a-th atom, $\Delta r_a = r_a(t) - \langle r_a(t) \rangle_t$ and $\langle \cdot \rangle_t$ represents the time average. This function returns a matrix of all atom-wise cross-correlations whose elements, $\mathrm{DCC}_{a,b}$, may be displayed in a graphical representation frequently termed as dynamical cross-correlation map, or DCCM.[132]

If $\mathrm{DCC}_{a,b}$ = 1 the fluctuations of atoms a and b are said to be completely correlated (same period and same phase), if $\mathrm{DCC}_{a,b}$ = -1 the fluctuations are said

to be completely anticorrelated (same period and opposite phase), and if $DCC_{a,b}$ = 0 the fluctuations of a and b are said to be not correlated. Typical characteristics of DCCMs include a line of strong cross-correlation along the diagonal, cross-correlations emanating from the diagonal, and off-diagonal cross-correlations. The high diagonal values occur where $a = b$, where $DCC_{a,b}$ is always equal to 1, (self-correlations). Correlations emanating from the diagonal indicate correlations between contiguous residues, typically within a secondary structure element or other tightly packed unit of structure. Finally, off-diagonal cross-correlations might be indicative of correlated motions between separate structural parts of a protein, e.g. between residues far away in terms of primary sequence. The DCCM analysis is usually performed on the C_α atoms.

3.4.5 Principal Component analysis

The Principal Component analysis (PCA)[133] is a powerful method used to analyze large collective motions in biomolecular systems. This method rests on separating the large concerted structural rearrangements from the less relevant fluctuations, in order to grasp the so-called "principal components" of the motion, i.e. the major collective modes of fluctuation that are assumed to dominate the functional dynamics. The power of PCA lies in the fact that usually a small number of modes suffice to describe the majority of the total motion of a system.

The analysis consists of two steps. First the configurations from the trajectory are superposed to enable filtering of the internal motions from overall rotations and translations, via least-squares fitting onto a reference structure. Then, this "fitted" trajectory is used to construct a covariance matrix of the structural positional fluctuations. This is a symmetric matrix containing, for each pair of atoms, the covariances (off-diagonal elements) and the variances (elements along the diagonal) of the atomic displacements, relative to their respective averages.

$$\mathbf{C} = \langle (\boldsymbol{x}(t) - \langle \boldsymbol{x} \rangle)(\boldsymbol{x}(t) - \langle \boldsymbol{x} \rangle)^T \rangle_t$$

Here the coordinates are provided as a function of time for clarity, but may be provided in any order. Atoms that move concertedly give rise to positive covariances, whereas anticorrelated displacements are associated with negative

covariances and non correlated displacements result in near-zero covariances. This symmetric matrix is subsequently diagonalized by an orthogonal coordinate transformation.

$$\mathbf{C} = \mathbf{T}\Lambda\mathbf{T}^{\mathbf{T}}$$

The diagonalization of this covariance matrix yields a set of eigenvectors, Λ, each associated with a collective mode, and corresponding eigenvalues, $\mathbf{T}$, which represent the relative contribution of each collective mode to the total variance of the system. The eigenvectors are usually sorted so that their eigenvalues are in decreasing order. If at least 3N configurations are used to construct $\mathbf{C}$, then 3N-6 eigenvectors with non-zero eigenvalues will be obtained. Indeed, six eigenvalues should be exactly zero, since the corresponding eigenvectors describe the overall translations and rotations (which are eliminated by superposition). The largest amplitude modes of a PCA, i.e. those characterized by the largest eigenvalues, usually also represent the slowest dynamical transitions of a system, i.e. the most functionally important.

Ultimately, once the eigenvectors have been obtained, the original trajectory can be further analyzed in terms of its principal components, a procedure better known as essential dynamics (ED) analysis.[133] Such analysis is carried out by projecting the trajectory onto each of the principal modes (usually the first two or three) to yield the time behaviour and distribution of each of the principal coordinates of the sampled distribution. PCA can be carried out on any subset of atoms and usually only C_α or backbone atoms are taken into account.

3.4.6 Electrostatic Potential Surface analysis

The electrostatic potential was calculated using the ABPS software.[134] The program takes as input a coordinate file of a molecule (usually a modified format of the PDB file including also atomic charges and radii) and calculates the electrostatic potential along its vdW or solvent accessible molecular surface by numerically solving Poisson-Boltzmann equations based on atomic point charges. The calculation was done using 2 and 80, as the solute and solvent dielectrics at 0.15 M salt concentration.

3.4.7 Pore Radius analysis

The analysis of pore dimension has been carried out by means of the HOLE program.[135] This program implements an algorithm based on a Monte Carlo simulated annealing procedure to find the best route for a sphere with variable radius to squeeze through a protein channel. The program require three inputs: 1) the coordinates of the protein of interest, 2) an initial point $\boldsymbol{z}$ which can lie anywhere in the pore and 3) a user specified vector $\boldsymbol{l}$ that is approximately along the direction of the pore (normally referred to as the channel direction vector). HOLE reads atoms from within the pdb file and sets a van der Waals radius for each. Then, the program looks for the maximum radius $R(\boldsymbol{z})$ of a sphere centred at point $\boldsymbol{z}$ without overlapping with the van der Waals surface of protein atoms:

$$R(\boldsymbol{z}) = \min^{N_{atom}} [|\boldsymbol{x}_i - \boldsymbol{z}| - vdW_i]$$

where $\boldsymbol{x}_i$ is the position of atom number i, vdW_i its van der Waals radius for atom i, and N_{atom} is the total number of atoms. By using the Monte Carlo simulated annealing procedure $\boldsymbol{z}$ is adjusted, while keeping the sphere centre on a plane normal to the channel direction vector, and the radius of the sphere maximized at a given position within the pore of the protein. Thus, the radius $R(\boldsymbol{z})$ can be regarded as an objective function of the point $\boldsymbol{z}$. The search is performed for a series of parallel planes of uniform separation along the pore, with a distinct optimization procedure being conducted for each plane. The net result can be thought of as producing the locus of a flexible sphere "squeezing" through the centre of the channel.

3.4.8 Protein Channels analysis

The identification of protein channels has been performed via the Molaxis webserver.[136] Given a protein structure in the pdb file format, the server outputs all possible channels that connect buried cavities to the outside of the protein. For each channel, the channel dimensions, the gating residues and the narrowest radius (termed as bottleneck) are also provided.

Molaxis is based on an algorithm that employs computational geometry techniques. The algorithm operates in several steps:

1. A representation of the input molecule using a collection of fixed size balls, whose number depends on the required quality of the approximation, is created.

2. A Voronoi diagram of the centers of the fixed-sized ball is computed.

3. A graph using the Voronoi vertices and edges is computed. To each edge is assigned a weight that depends on its length and on its distance from the van der Waals surface of the molecule.

4. A minimal weight three from a user-specified source point or from an starting point automatically defined as the centre of the largest chamber in the molecule is computed, by means of a shortest-path optimization algorithm.

5. The so-constructed paths or corridors, each representing a probable route taken by a small molecule passing through a channel, are scored according to the total weight of the edges in the graph.

Chapter 4

The 5-HT_3A receptor

In recent years, X-ray crystallographic structures of the pentameric ligand-gated ion channels (pLGICs) have boosted the research in computational biology, owing to their importance in vital neuronal activities and their suitability as targets for pharmacological treatments. pLGICs form the so-called Cys-Loop receptor superfamily and are involved in a number of different physiological and pathological roles[30]: in signal transduction processes at synaptic level,[137] pLGICs can change membrane potential by allowing ions to transiently translocate through them via a complex (not fully uncovered) gating mechanism triggered by neurotransmitter binding. Members of this family are integral oligomeric membrane proteins made up by three domains [33,138]: I) an extracellular domain (ECD), which contains the binding pocket of the neurotransmitter (at the interface of two subunits), II) a transmembrane domain (TMD) characterized by a bundle of α-helices spanning across the membrane and forming the channel core region and III) an intracellular domain (ICD), which has a role in the localization and regulation of the receptor and whose size could vary among the members of the family (Figure 4.1A). Since the discovery of the first member of the family, the nicotinic acetylcholine receptor (nAChR), these receptor channels have been intensively studied through a considerable number of experimental [139-142] and, recently, also computational [143-149] approaches. To date the residues forming the binding pocket have been identified,[49,150] as well as those responsible for the channel gating upon agonist binding.[139,141-145,147-149,151-162] Also residues affecting biophysical properties, like ion selectivity and conductance, and pharmacological properties have been

discovered.[140,146,163-168] Despite the progress towards the full comprehension of pLGICs, many issues are still unsolved primarily because of scarce information on the conductive channel structures. Besides, it is still unclear the relation between the observed holo-open (conductive) and holo-desensitized (non-conductive) states. As a consequence, most experiments and computational studies rely on the resolved closed-state structures for the interpretation of many pLGICs properties.

Among others, the 5-Hydroxytryptamine receptor type 3 (5-HT_3R), which is widely expressed throughout the brain and in peripheral organs (e.g., the gastrointestinal tract), has been extensively studied from a structural and functional point of view.[169] The 5-HT_3R is involved in many neurophysiological functions,[170,171] like nervous transmission regulation, pain processing, peristalsis and in some pathological states[171] like chemotherapy-induced nausea and vomiting and various psychiatric disorders. Five different subunits of this receptor have been identified in the genome: among these, only receptors formed by subunits A and B have been extensively investigated and characterized in mammals, while the biological relevance of other forms made up by subunits C, D, and E has not been fully elucidated. From a biophysical point of view, the homo-pentameric 5-HT_3A receptor is by far the most studied form, due to the capability of subunit A to assemble into functional homomeric channels. Successful experimental efforts on the 5-HT_3A receptor led to the identification of the residues shaping the binding site,[172-180] the inner surface of the M2 helices[181], which circumscribe the transmembrane portion of the channel, and the membrane associated (MA) helices in the ICD.[182] Moreover, an extensive work has been carried out in order to unravel the nature of the amino acids coupling the agonist binding to the channel gating,[139,142,155,158] as well as those responsible for the cation-selectivity[183,184]. Also, residues responsible for the very low single-channel conductance, which differentiate the 5-HT_3A receptor from other members of the family, have been identified.[185] However, despite all these efforts, many structural and mechanistic details of 5-HT_3A still remain poorly understood. In the present work, taking advantage of the recently resolved mouse 5-HT_3A crystal structure in a putative closed state,[48] we have carried out the first computational study of such a receptor channel to investigate protein dynamics as well as ion-receptor interactions. In particular, through the use of atomistic molecular dynamics simulations and free energy calculations, we have studied three aspects of the receptor: (1) the structural and dynamical

features of the whole channel, including ECD, TMD and part of the ICD (MA helices); (2) the potential of mean force of the single-ion translocation through TMD and ICD (the latter formed by post-M3 loops and MA helices) in order to identify and evaluate the energetic barriers experienced by the ions crossing regions pivotal for ionic conductance and selectivity; (3) the molecular determinants of the low single-channel conductance. In the latter case, we have simulated three systems, i.e. a wild-type protein and two high-conductive mutants (namely, the R420D single mutant and the R416Q/R420D/R424A triple mutant, according to the crystal structure numbering adopted in this work). Our overall ~1 μs simulation data provided evidences for the existence of intersubunit lateral pathways at the ECD-TMD interface used by ions to access the channel interior and suggested the possible influence of residues, not investigated so far, on single-channel conductance and cation-selectivity. Moreover, peculiarities, similarities and differences in structural and biophysical properties, as compared to other well-known homologue channels, are thoroughly presented and discussed.

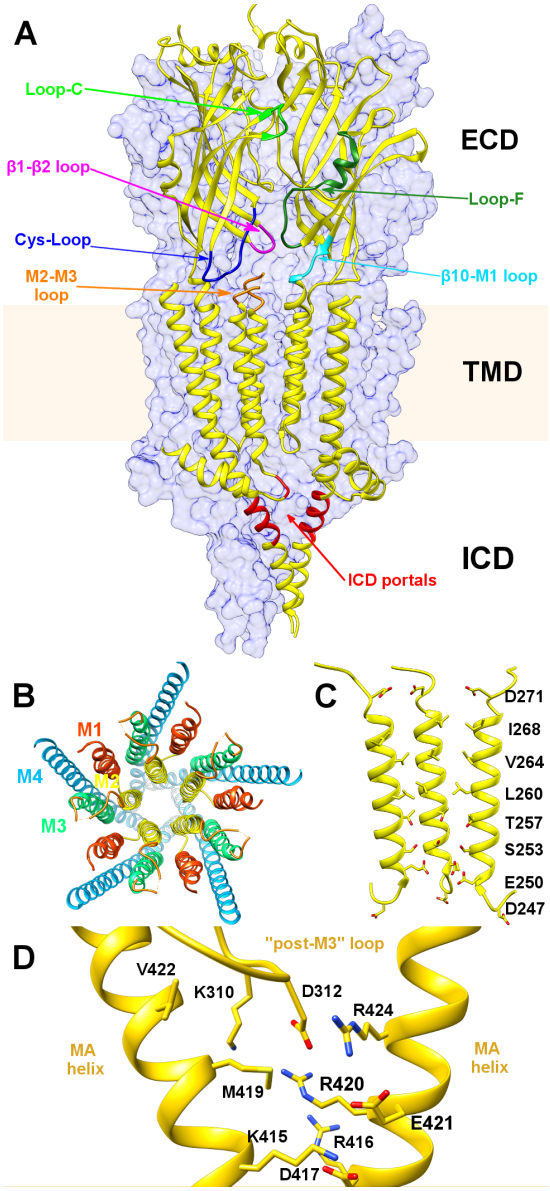


Figure 4.1: **Topology of the murine serotonin 5-HT$_3$A receptor (PDB entry: 4PIR).** (A) Side view of the homo-pentameric receptor along the channel axis, showing the three main domains (ECD, TMD and ICD). The two front subunits are shown as yellow cartoons. The remaining three subunits are shown as surfaces colored in blue. (B) Top view of the TMD helices shown as ribbons. (C) Side view of the M2 helices and residues that line the channel interior, with side chains displayed as sticks. The two front helices have been removed for clarity. (D) Enlarged view of the residues shaping the ICD portals, shown as sticks.

4.1 Computational details

4.1.1 Systems preparation

Whole 5-HT$_3$AR channel. The 5-HT$_3$AR channel (PDB entry: 4PIR) was first oriented and then inserted into a POPC (1-Palmitoyl-2-Oleoyl-sn-Glycero-3-Phosphatidylcholine) lipid bilayer using the OPM[186] and CHAMM-GUI[187] servers, respectively. The protein-lipid assembly was then solvated with TIP3P water molecules and Na^+ and Cl^- ions to neutralize the overall charge and obtain a final ion concentration of about 0.15 M, using the VMD[188] software. Ions were introduced at least 8 Å away from the protein. Since a long part of the M3-M4 loop was absent from the structure, the C-terminus of the post-M3 segment and the N-terminus of the MA helix were capped with a methylamine and an acetyl groups, respectively, so as to mimic a more realistic situation where these segment termini are neutral. The M2-M3 loop segment that was unresolved in the crystal structure has been reconstructed using the Modeller program (v9.12).[189] All histidines were set in neutral form, since none of them was found in a context such to justify a protonated state. The resulting simulation box size was 154 Å × 154 Å × 214 Å and the total system size about 400,000 atoms .The system was then minimized in two steps: first, a 20000 step minimization run was carried out with restraints on all protein atoms (5 kcal/mol/Å^2); then, a further 15000 step minimization was carried out with 1 kcal/mol/Å^2 restraints on C_α atoms only. Afterwards, the system was slowly heated up from 1 to 310 K over a 4 ns MD simulation, applying gradually decreasing restraints on the protein C_α atoms, from 5 to 1 kcal/mol/Å^2. After such thermalization, the system was equilibrated for further 2 ns, gradually reducing to zero the atomic restraints on the protein flexible parts (ECD, loops, ICD).

Truncated wild-type and mutant channels (TMD+ICD). The procedure to prepare the smaller systems made up by TMD and ICD was slightly different. We first created the mutants starting from the full 5-HT$_3$A crystal structure (PDB entry: 4PIR). The three systems (WT, R420D and QDA) were then embedded into the lipid membrane, solvated, thermalized and equilibrated as described above. After equilibration, we created a reduced system (starting at P213, in the β10-M1

loop) removing the ECD from protein structure, capped P213 N-terminus and adjusted the box volume in order to ensure, at least, a 20 Å thick water layer. Ionic concentration was set again to 0.15 M. Special attention was paid to avoid ion inclusions inside the ICD vestibule. The resulting simulation cell was 156 Å x 154 Å x 148 Å, with about 260,000 atoms. Finally, the three systems were further equilibrated for 2 ns prior to production runs.

4.1.2 Simulation and analysis details

Molecular dynamics. All MD simulations were performed using the NAMD v2.9 program.[190] Protein, water and ions were modeled with the CHARMM27 force field[191] with CMAP corrections for backbone atoms and NBFIX correction for Na^+ ions,[192,193] while the CHARMM36 force field[194] was used for lipids. On the C_α atoms of only TMD helices, soft symmetry restraints available in NAMD, with a force constant of 0.75 kcal/mol/Å^2, were applied. ECD, ICD and other more flexible parts of the protein (e.g. loops) were kept unrestrained. It is worth noting that these restraints do not cancel out global movements of the TMD, but rather ensure, on average, a symmetrical homopentameric assembly, as expected in the present system. Simulations were performed under constant pressure (1 atm) and temperature (310 K), where the NPT ensemble was enforced by the Langevin piston pressure control and the Langevin damping dynamics using PBC. All covalent bonds with hydrogen atoms were kept rigid using the SHAKE algorithm. The bonded interactions and the short-range non-bonded interactions were calculated at each time-step (2 fs), whereas the PME method was used to update the long-range electrostatic interactions every two time-steps. PMEGridSpacing was set 1.0 Å. The cutoff distance for non-bonded interactions was 10 Å. A smoothing function was employed for the van der Waals interactions at a distance of 8 Å. The pair list of non-bonded interactions was evaluated using a pair distance of 14 Å. The whole receptor was simulated for a total time of 100 ns. WT, R420D and QDA systems were simulated for 100 ns each, applying symmetry restraints.

Free energy calculations. From the crystal structure of the whole 5-HT_3A receptor, a smaller subsystem comprising only TMD and ICD was created. Only the ions needed to ensure the system electroneutrality were introduced. After 2 ns equilibration, we performed a steered MD simulation, slowly dragging one Na^+

ion (or one Cl^-) through the pore in order to generate multiple starting configurations suitable for the ABF windows. Each ion was dragged along the TMD pore in about 30 ns. Subsequently, we equilibrated each window for 2 ns prior to free energy calculations. The ABF method as implemented in NAMD[122,126,128], was used to calculate the single-ion PMF for Na^+ and Cl^- along a reaction coordinate spanning ~65 Å, sufficient to cover the entire length of the TMD and the upper part of the ICD. The simulations were carried out in 1 window of 3 Å length and 11 windows of 5 Å length along the channel axis. The same simulation conditions described above were applied for this calculation. Within each window the average force acting on the selected ion was accumulated in 0.1 Å sized bins, and the biasing force was applied after 800 samples. Positional restraints (with force constant of 3 kcal/mol/$Å^2$) were applied to counter-ions so to keep them away from the ion inside the channel and rule out any interference in PMF calculation. A 20 ns MD trajectory was generated for each window, resulting in a total simulation time of 240 ns per ion.

Trajectory analysis. The HOLE[195] program and the MolAxis[136] webserver were used to analyze pore dimension and fenestrations. RMSF and B-factors were calculated using the ptraj[131] software tool. RMSD, ion density and coordination analyses were performed using in-house scripts exploiting the MDAnalysis[196] library. Electrostatic potentials were evaluated through the APBS[134] server. Figures and plots were generated using the UCSF-CHIMERA[197] and MATPLOTLIB[198] software.

4.2 Results

4.2.1 Structural features of the 5-HT3A receptor

In Figure 4.1A, the structure of the 5-HT_3A receptor is depicted, where the three domains forming the integral homopentameric membrane protein are evidenced: I) the ECD is mostly made up by ten β-sheet strands along with the corresponding loops interconnecting them; II) the TMD is formed by an α-helix bundle of four α-helices per monomer, namely M1, M2, M3 and M4, spanning the lipid bilayer, where the assembly of the internal M2 helices frames the channel core (Figure 4.1B). The pore walls are characterized, in the central region, by a

stretch of hydrophobic residues, while polar and charged residues are located at both ends of the TMD channel (Figure 4.1C). In particular, it is believed that the gating mechanism, which is triggered by agonist binding, does involve a rather complex series of movements leading to the pore expansion/compression at the TMD level (see, for example, ref.[151] for a recent account of the gating mechanism in pLGICs); III) the ICD is formed by an α-helix (MA-helix) and a long, partially unresolved, loop connected to the M3 helix in the TMD. The entire 5-HT_3A receptor was simulated for about 100 ns, without applying an external electric field, after being embedded into a lipid membrane and solvated by adding water and 0.15 M NaCl (details are provided in the Methods section).

First, we evaluated the system RMSD with respect to its starting molecular configuration in order to test the system stability throughout the simulated time interval (Figure 4.2).

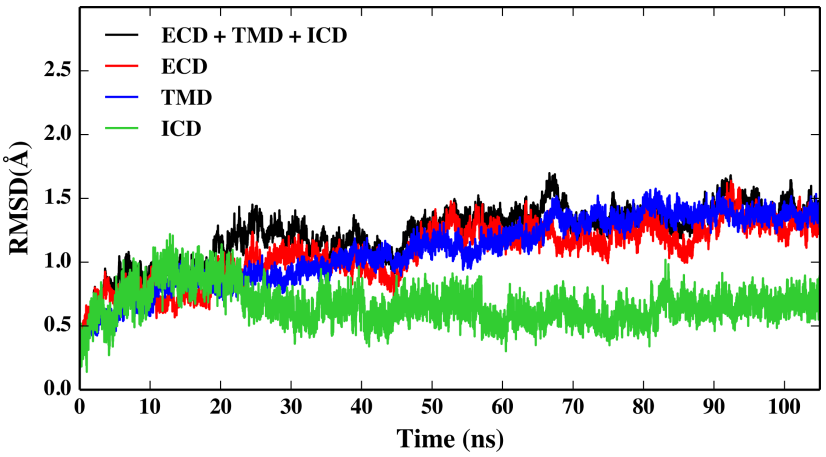


Figure 4.2: Backbone RMSD with respect to the starting frame.

The RMSD for all the three domains of the receptor, as analyzed separately, was stable (below 1.5 Å) during the entire simulation, hence the system maintained a stable pentameric assembly over time.

Afterwards, we analyzed the main protein structural fluctuations through the evaluation of the RMSFs. The RMSF analysis revealed that the protein most flexible regions are the loops C and F in the ECD and the M2-M3 loop, as shown in Figure 4.3, while the rest of the protein appeared, overall, quite rigid.

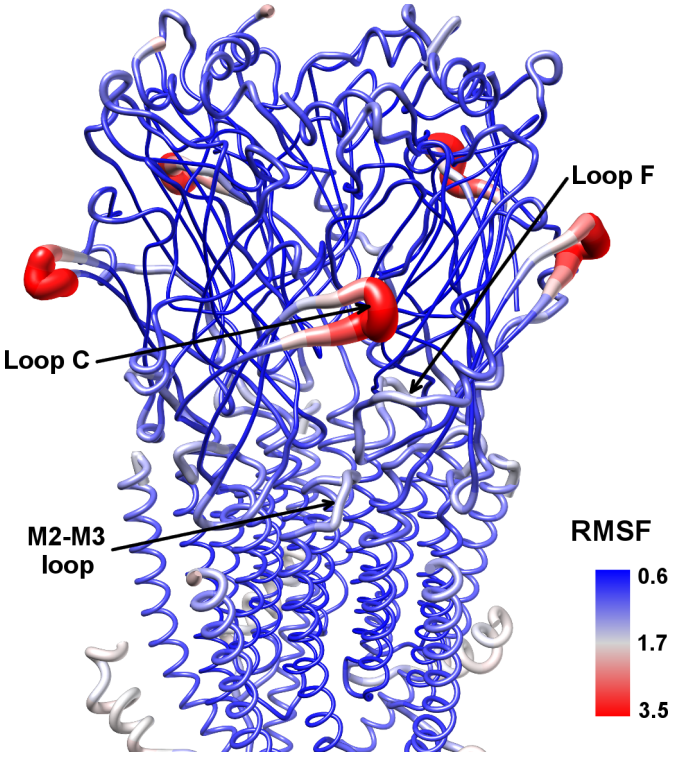


Figure 4.3: Backbone RMSF. The protein average structure, depicted as ribbons, is colored according to average backbone RMSF values, from lowest (blue) to highest (red). The thickness of the ribbons is also proportional to RMSF values.

We also analyzed the channel dimension along the receptor longitudinal axis (i.e., *z*-axis), in terms of the pore radius as provided by the HOLE program, and found a funnel-like profile with a minimum constriction point located in the TMD (Figure 4.4A), already pinpointed as the location of the channel gate[143,181]. Note that a first constriction point was observed at position K108 (at $z = \sim -30$ Å) in the ECD. Here, the obtained pore radius is around 3.8 Å, which makes this site almost as constricted as the upper part of the hydrophobic segment in the TMD (at $z = \sim 5$ Å). In the crystal, a sulphate ion from the crystallization liquor was found to interact with the ring of charged K108 residues, hence the K108 side chains adopted an extended conformation towards the centre of the channel, considerably shrinking the pore size at this site. In our simulation, the K108 side chains were folded back to interact with the D105 side chain of adjacent subunits (Figure 4.5).

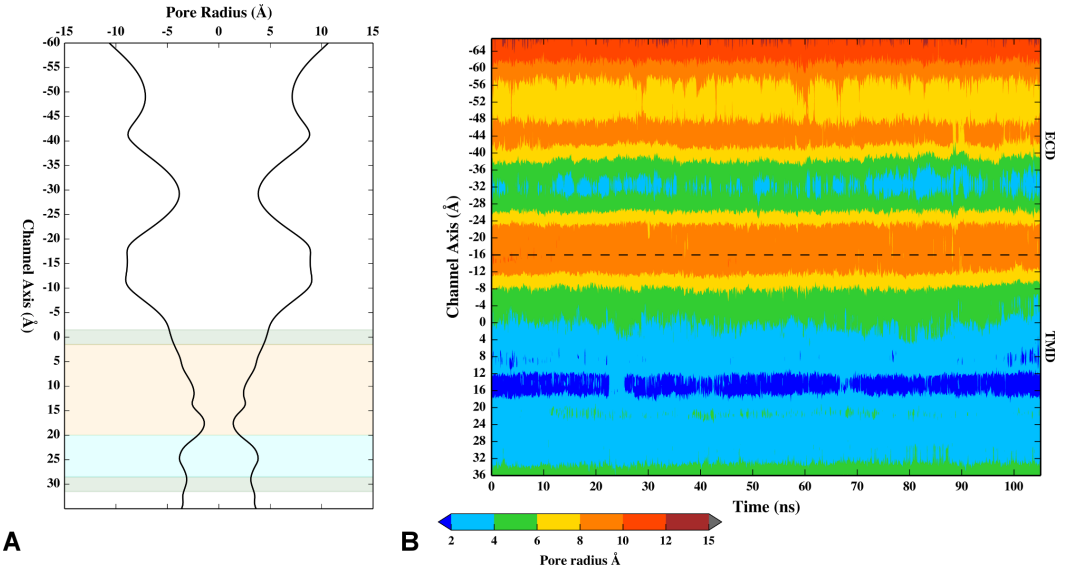


Figure 4.4: **Pore dimension analysis.** (A) Average pore radius as a function of channel axis position (mirror symmetry applied for clarity). The shaded bands highlight different regions of the TMD, namely negatively charged (green), hydrophobic (orange) and polar (cyan) residues. (B) Time series of the pore radius as a function of channel axis position. The dotted line identifies the interface between ECD and TMD. In both graphs the zero is taken as the centroid of D271 C_α atoms. Negative values represent the extracellular side of the pore, while positive values represent the intracellular side.

In the channel hydrophobic region the pore radius at residue V264 and L260 reached its minimum (3.0 Å at V264 and 2.5 Å at L260), making L260 the narrowest point along the TMD. Note that the time evolution of the pore dimension has shown only little deviations from the average structure (Figure 4.4B) and the initial crystal structure (i.e., pore radius ~3.3 Å at V264 and ~2.3 Å at L260) during the dynamics, supporting an apparent closed state of the 5-HT_3A channel (i.e., not compatible with the passage of aqua ion clusters). Albeit in the crystal structure the maximum constriction point was reported further down (z = ~ 30 Å) at position E250 (i.e., pore radius ~2 Å), the side chains of these residues were not well resolved and were modelled as most plausible rotamers. Therefore, the reported pore size estimate has to be considered rather uncertain. From our MD simulation, we have obtained, on average, a pore radius of about 3.3 Å at

residue E250, which is perhaps a more appropriate estimate considering the effect of side chain repulsion.

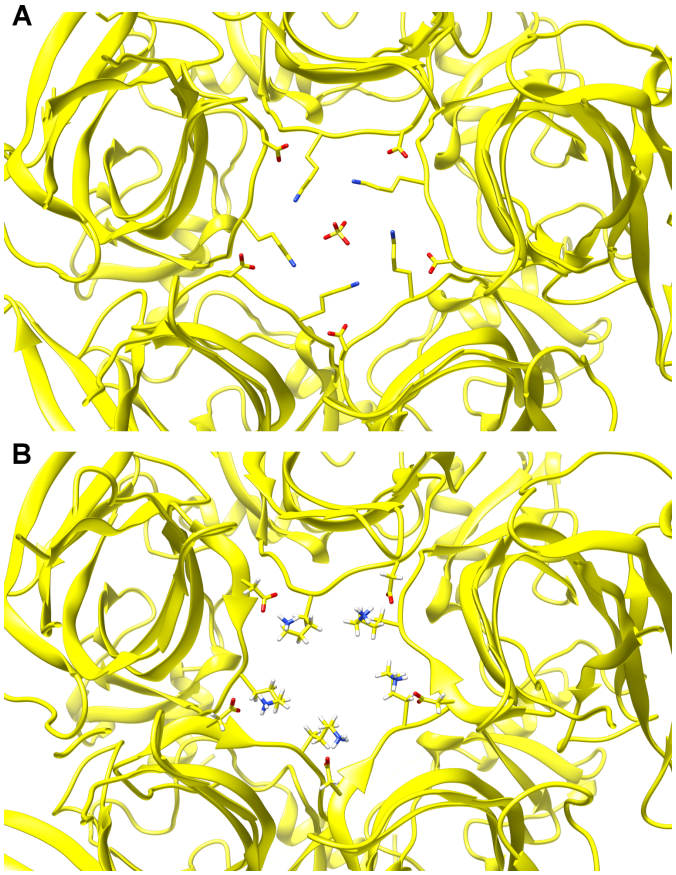


Figure 4.5: K108 Side chains conformation. Bottom view of K108 side chain conformations in the (A) 5-HT_3A crystal structure and (B) a representative snapshot from the simulation. The protein is depicted as yellow ribbons, while the side chains of K108 and D105 are depicted as sticks.

4.2.2 Ion permeation into the receptor body

The ion permeation through the 5-HT_3A receptor was evaluated by considering an initial configuration with no ions within the entire protein body and then by monitoring the ion entrance through the accessible pathways. A cylindrical region of 25 Å radius aligned along the channel axis, including the whole ECD, TMD and part of the ICD, was considered in order to follow the

permeation events. Note that ion permeation was observed only from the extracellular side of the receptor. Figure 4.6 depicts the distribution of ion (Na^+ and Cl^-) positions sampled during the MD simulation.

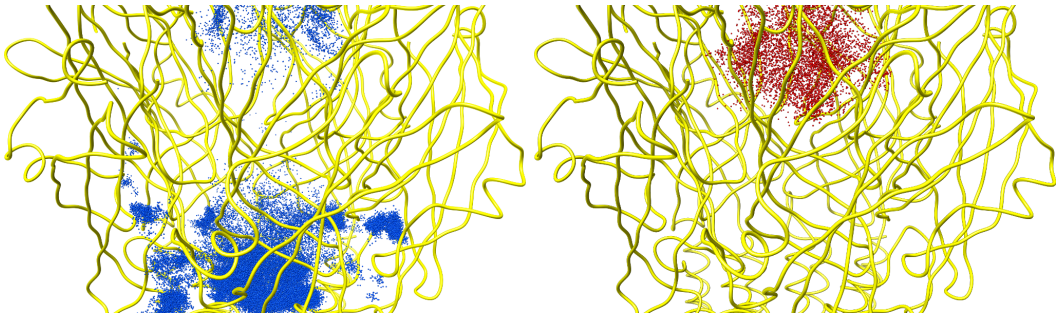


Figure 4.6: **Ion distribution inside the 5-HT$_3$A channel.** Na^+ (blue) and Cl^- (red) ion positions, as issuing from the MD the simulation, sampled inside and outside the receptor (a cylinder of 25 Å radius has been defined, taking the centroid of D271 C_α atoms as the center). Note, at the bottom of left panel, the distribution of Na^+ ions permeated into the receptor through the ECD lateral pathways. The average structure of the protein is shown as yellow ribbons.

It is apparent that the presence of Cl^- inside the ECD vestibule of the channel is essentially smaller than that of Na^+, though not negligible. Interestingly, we have observed that the top side of the channel is not the only way through which ions may access the channel ECD vestibule. Indeed, cations can also enter into the ECD passing through five lateral channels located at the interface between the ECD and TMD of any pair of adjacent subunits. Such channels are shaped by residues belonging to Cys-Loop, β1-β2 loop, M2-M3 loop, β10-M1 loop and the C-terminal part of the F-loop: to be specific, the lateral channels are formed by residue D52, E53, K54 and N55 from the β1-β2 loop, by residue Q184, G185, E186 and, to a minor extent, E188 from the F-loop C-terminal part, by residue R219, P220, L221, F222 and, to a minor extent, R217 and R218 from the β10-M1 loop, by residue C135, S136, L137, D138 from the Cys-Loop and a stretch of residues (i.e., P274-T280) from the M2-M3 loop (Figure 4.7).

From our MD simulation, such lateral pathways appeared to be the main entering route to the ECD vestibule: we counted at least 5 different permeating Na^+ ions passing through the lateral channels against one Na^+ ion entering from the channel top side (note that we have observed Na^+ ions passing through at least

three of the five available lateral channels, which are all equivalent due to symmetry). The presence of only one top entryway versus five equivalent lateral ones, though featuring a smaller diameter, may suggest a detectable contribution of the latter to the overall single channel conductance. In order to better characterize the size of such lateral cavities, we submitted an average receptor configuration issuing from the MD simulation to the MolAxis webserver. Figure 4.8 displays the lateral channels and their corresponding diameter size, as provided by the MolAxis analysis tool.

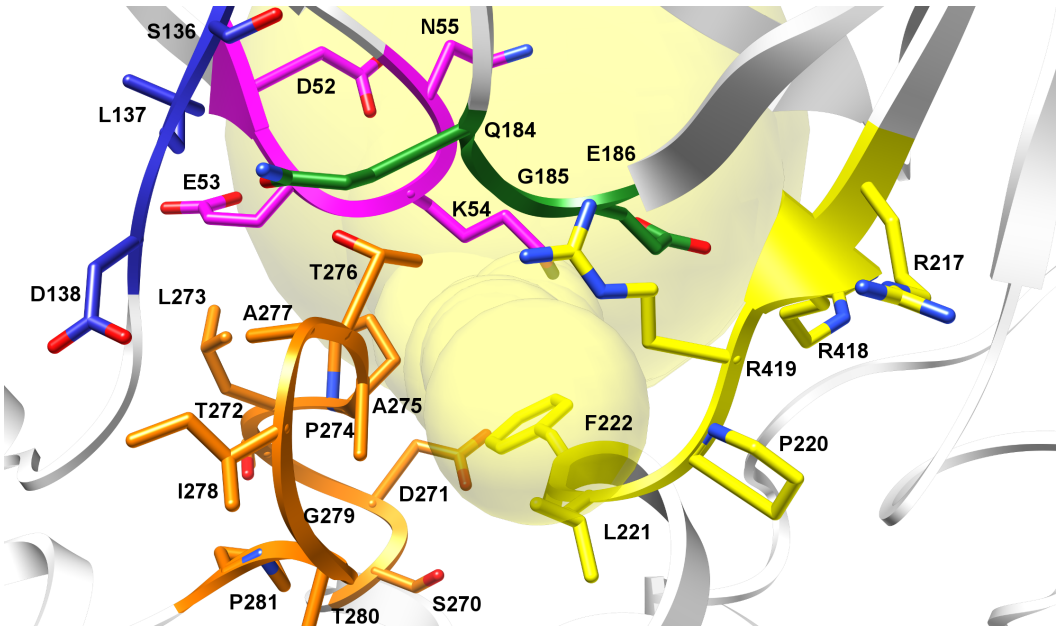


Figure 4.7: Enlarged view of a lateral portal at the ECD-TMD interface. The protein structure is depicted as cartoons, with residue side chains molding the portal walls as sticks. Cys-Loop, β1- β2 loop, C-terminal part of the F-loop, M2-M3 loop and β10-M1 loop are highlighted in blue, magenta, green, yellow and orange, respectively. In background, one channel predicted by MolAxis is depicted as a series of transparent yellow spheres, whose radius is proportional to the accessible space.

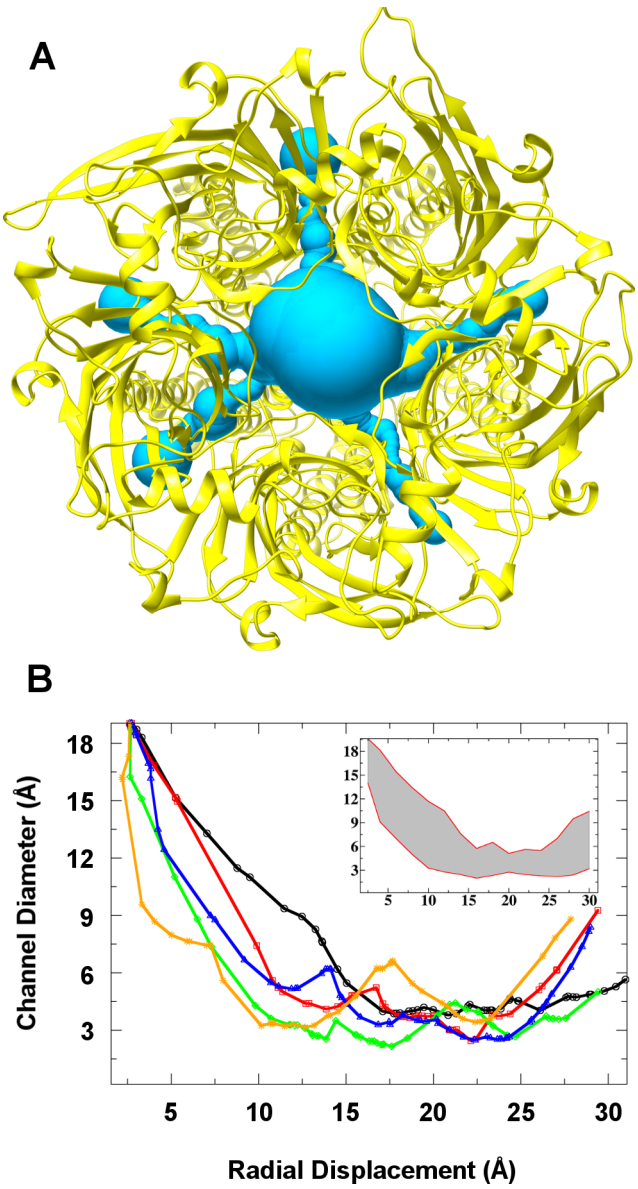


Figure 4.8: Intersubunit lateral ion pathways. (A) Top view of the 5-HT_3A receptor principal and lateral channels predicted by MolAxis, shown as a series of blue spheres with diameter proportional to the accessible space, and (B) the corresponding lateral pore dimensions (the inset displays the maximum channel diameter deviations from the average (grey band) as issuing from the analysis of several protein structures). The protein average structure is depicted in yellow cartoons.

Since the maximum constriction point lies below the diameter of Na^+ first solvation shell (about 6 Å), we concluded that these channels are not constitutively open, but may transiently get expanded to allow ion passage (see inset of Figure 4.8). Motivated by the present results, we wondered if such structural features could be common to other 5-HT_3A receptor analogues. Similarly, we analyzed the known structure of nAchR (PDB entry: 2BG9)[199], GluCl (PDB entry: 4TNV)[200], ELIC (PDB entry: 3RQU)[201], GLIC (PDB entry: 3P50)[202] and finally 5-HT_3A crystal structure (PDB entry: 4PIR) as a reference, after modelling the missing part of the M2-M3 loop. Surprisingly, MolAxis predicted the existence of such channels not only for the 5-HT_3A crystal structure but also for nAchR and GluCl, whereas they have not been found in GLIC and ELIC (Figure 4.9).

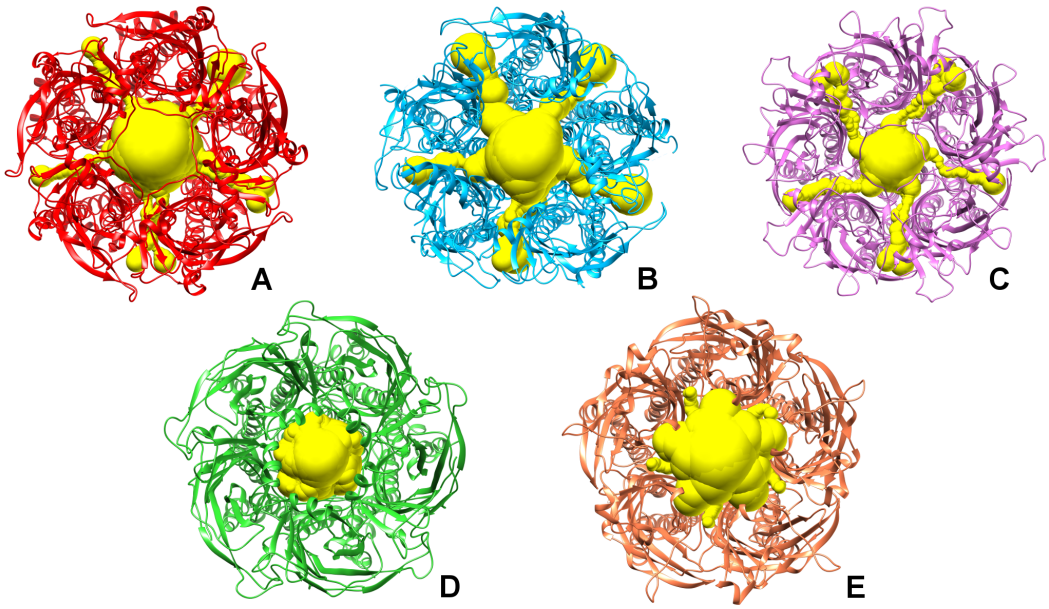


Figure 4.9: Lateral channels in other members of the Cys-Loop superfamily. Lateral intersubunit pathways from the interior to the exterior of the protein in: (A) 5-HT_3A, (B) nAchR, (C) GluCl, (D) ELIC, (E) GLIC.

Then, we analyzed the ion distribution along the 5-HT_3A channel axis in terms of the average occurrence number (Figure 4.10). Results have shown an extended region roughly spanning 35 Å along the pore axis, where neither chloride nor sodium ions were found; this region corresponds to the TMD central

region, comprising both a purely hydrophobic (from I268 to L260 of M2 helix) and a polar (from T257 to E250 of M2 helix) stretch, and the first part of the ICD. At the beginning of the TMD (around z = 0 Å), it has been observed a spike in the average number of Na^+, while the presence of Cl^- was basically null. Here, just before the hydrophobic region characterizing the TMD pore, a ring of aspartates (D271) wards a funnel-shaped zone which seems well suited to act as a pool for concentrating cations before their subsequent translocation along the channel, as already proposed in previous studies on cation selective channels.[146,203,204] In light of these results, we decided to investigate in more detail the role played by the TMD in hindering the ion passage in the closed state and in selectively filtering cations versus anions.

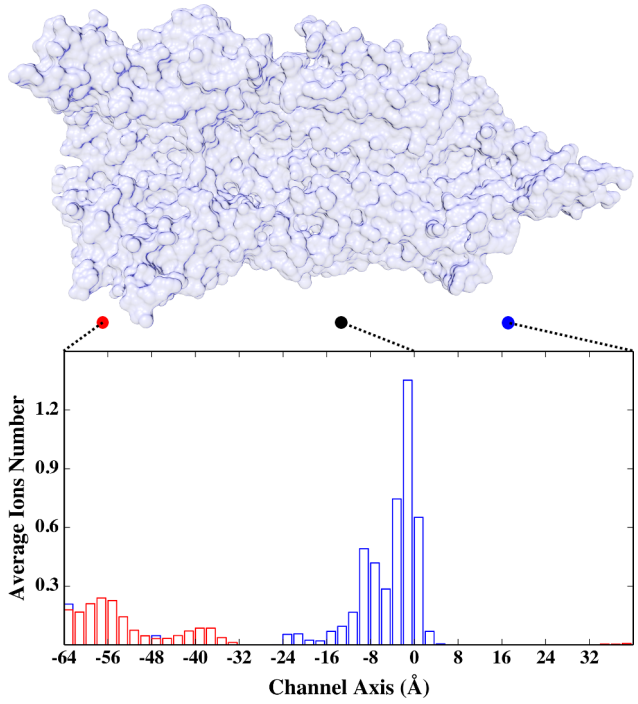


Figure 4.10: Average number of ions along the channel axis. Bar histogram of the average number of Na^+ (blue) and Cl^- (red) ions along the channel axis (inside a cylinder of 25 Å radius). Negative values represent the extracellular side of the pore, while positive values represent the intracellular side. Origin is set to the centroid of D271 C_α atoms.

4.2.3 Free energy of ion translocation through the TMD

In order to investigate the role of TMD residues on ion transport properties, we evaluated the single-ion PMF of both Na^+ and Cl^- inside the TMD pore.. The present analysis is undoubtedly limited by the availability of only the closed state 5-HT_3A structure. However, as shown in several previous studies on pLGICs closed-state channels,[143,146,203] a general qualitative picture of the pore can be satisfactorily obtained, including some important features such as a molecular description of the selectivity filter. Furthermore, we complemented such an energy analysis with a detailed picture of the variable ion coordination number along the pore, evaluating the average number of water molecules and side-chain atoms in the first solvation shell. To this end, a subsystem of the whole receptor protein made up by only the TMD and ICD was obtained from the mouse 5-HT_3A crystal structure and embedded into lipid membrane and water. Apart from a few ions ensuring the system electric neutrality, only one ion (either Na^+ or Cl^-) was inserted in the TMD for carrying out the free energy analysis and the resulting system simulated for an overall time of about 240 ns (see Methods for details). Figure 4.11 shows the obtained PMF profiles for both considered ions.

At the extracellular end of the TMD, both Na^+ and Cl^- experience a set of negatively charged residues, namely D271 at position $z \sim -10$ Å along the channel axis. The Na^+ PMF shows a pronounced minimum in this region (~ -9 kcal), which is also characterized by a relatively wide pore diameter (~9.5 Å). This result supports the hypothesis that this zone may serve as a funnel for cations,[146] increasing their local concentration before permeating into the narrowest portion of the channel. Conversely, the marked increase of the Cl^- PMF (~ 9 kcal) in the same region indicates that the negative residues effectively do repel anions from the pore and thus disfavor their further translocation through the channel. Below residue D271, the M2 helices expose hydrophobic side chains (I268, V264, L260) towards the interior of the pore. The PMF in this region (from z = -5 to +10 Å along the channel axis) reaches a local maximum for both ions in correspondence of V264 (~ 5 kcal and ~14 kcal for Na^+ and Cl^- , respectively, around position $z = \sim 0$ Å along the channel axis). At this channel location, both Na^+ and Cl^- lose, on average, around one water molecule from their first solvation shell without compensating this loss with side chains interactions. These results corroborate the hypothesis regarding the hydrophobic nature of the gate,

previously proposed on the basis of experimental and computational studies on homologue receptors,[143,205] and are in qualitative agreement with previous PMF calculations performed on the α7 acetylcholine receptor.[146]

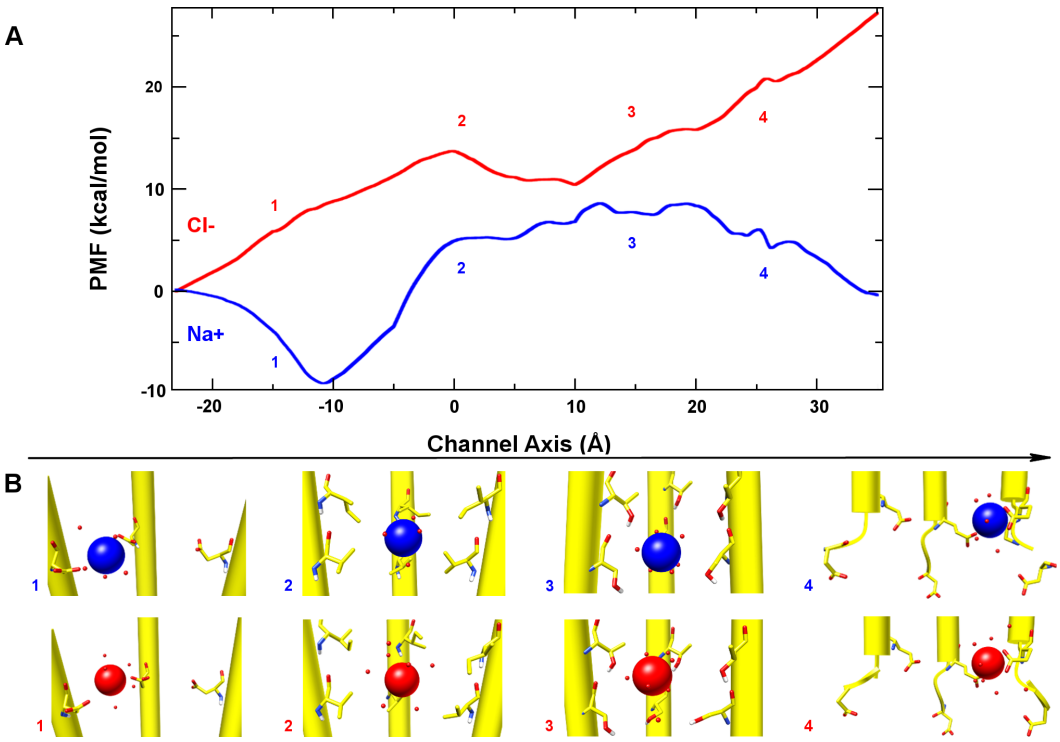


Figure 4.11: **Free energy barriers to ion translocation**. (A) PMF profiles for Na^+ and Cl^- along the channel axis. Channel axis origin is set to the centroid of V264 C_α atoms, at the center of TMD. Negative values refer to the extracellular side, while positive values refer to the intracellular side. The numbers (1-4) in the plot indicate: (1) negative, (2) hydrophobic, (3) polar and (4) negative regions of the pore, as described in the text. (B) Representative snapshots of Na^+ (blue) and Cl^- (red) along the TMD pore.

In the polar region of the TMD, lined between residue T257 and S253, (from z = +10 to +20 Å along the channel axis) the PMF profile is little affected in the case of Na^+, while it keeps increasing for Cl^-. A careful inspection at the ion coordination profile (Figure 4.12) shows that in the case of Na^+ the average number of coordinating species remained about 5 (i.e., the loss of water molecules is compensated by the close interaction with T257 and S253 side chains), thus

supporting the rather flat PMF observed in this region. Conversely, Cl^- has lost more water molecules than acquired polar side chains in its first coordination shell and, in turn, this led to a PMF increase while going towards the ICD. Note that the free energy of ion translocation in the present narrow pore segment could be also related to the difference in first solvation shell size between Na^+ and Cl^-.

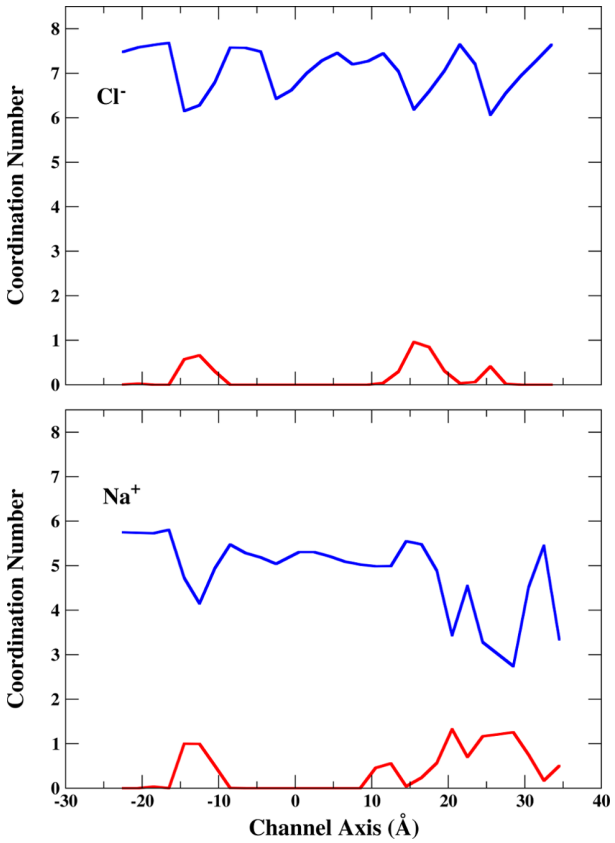


Figure 4.12: Average ion coordination number along the channel axis. The blue and red curves represent the average number of water and protein side chains atoms, respectively. The values are obtained from the last 5 ns of the ABF simulations.

To proceed further, after residue S253, the pore presents two rings of negatively charged residues, namely E250 and D247, located at the end of the M2 helix and on the M1-M2 loop, respectively (from z = +20 to +35 Å along the channel axis). In previous studies on the *Torpedo* nAchR and other homologue protein models built up from this structure,[143,146] the interpretation of the

computed single-ion PMF at the intracellular end of the TMD turned out to be problematic, owing to the poor structural resolution of the nAchR channel at this site and to the lack of the M3-M4 loop, which does interact with the M1-M2 loop affecting the conformation of the latter. On the other hand, the 5-HT_3A crystal structure presents both a higher atomic resolution and more complete portion of the M3-M4 loop. In turn, this allowed us to obtain a more accurate estimate of the PMF, as compared to previous studies. In this region of the pore, Na^+ and Cl^- PMF profiles reflected opposite electrostatic interactions with E250 and D247. On one hand, these interactions further destabilize the passage of Cl^- ions inside the pore, and on the other hand they stabilize significantly Na^+ ions. Accordingly, at this site the two PMFs show the maximal deviation: not surprisingly, residue E250 has been identified as the main selectivity filter of the 5-HT_3A channel based on conductive state experiments. Our findings also suggest an important role for D247, especially for providing the exit driving force for Na^+ ions from the TMD, and further considerations are provided in the Discussion.

4.2.4 Mutation effects on the ICD portals

In a seminal study on the human 5-HT_3A receptor,[185,206] whose results have been later confirmed also on the mouse ortholog,[207,208] it has been shown that the replacement of an arginine triplet (R416, R420, R424) located in the MA helix with corresponding residues from the 5-HT_3B subunit markedly affected the single-channel conductance. This evidence suggested that such mutations could facilitate the ion passage through lateral portals formed by adjacent MA helices,[206] otherwise hindered in the wild-type protein. These lateral openings are molded by a small stretch of residues belonging to the MA helices, i.e. 415-424, as identified by previous functional studies.[182] In particular, the portals are lined by residue K415, R416, M419 and V422 from one subunit, and residue D417, E421, R420 and R424 from the adjacent one (Figure 4.1D). The “ceiling” of the portals is formed by the starting portion of the M3-M4 segment (henceforth referred to as the “post-M3” loop, following the nomenclature reported in ref.[48]), especially by residue K310 and D312. Hydrogen-bond and salt-bridge networks among such residues apparently make the ICD cavities closed in the native receptor. In this study, we investigated in some detail the ICD structure of the mouse 5-HT_3A channel, which is

characterized by an unusual large fraction of well-resolved atomic coordinates, in order to better describe the proposed hampering effect towards ion permeation modulated by the present arginine triplet. Inspired by the findings of the mutagenesis experiment mentioned above (and in the absence of known structures of the mutants probed in that study), three model 5-HT_3A receptor systems made up solely by TMD and ICD were created, namely one wild-type (WT), one single mutant (R420D) and one triple mutant (R416Q/R420D/R424A, hereafter referred to as QDA). All systems were simulated for about 100 ns under physico-chemical conditions mimicking the natural environment (see Methods for details). Although part of the M3-M4 loop is truncated in the 5-HT_3A crystal structure, the present ICD structure represents the most complete model for studying ion diffusion through intracellular portals. Figure 4.13 shows the distributions of Na^+ ion positions, sampled during the simulations, inside and outside the ICD for the three considered systems.

In WT, no ions have been observed to pass from the aqueous solution to the intracellular vestibule; in the R420D mutant, one ion was able to translocate through the portals in the considered time interval and, in QDA, ions freely diffused from the solution into the ICD. To test if such differences in ion diffusivity were the result of structural changes or differences in the domain dynamics upon mutation, we calculated the RMSD and the per-residue RMSF on backbone atoms, especially focusing on those residues lining the portals walls, as described above. The RMSD for the three systems was essentially indistinguishable, with differences of less than 0.4 Å and overall similar broadness (Figure 4.14).

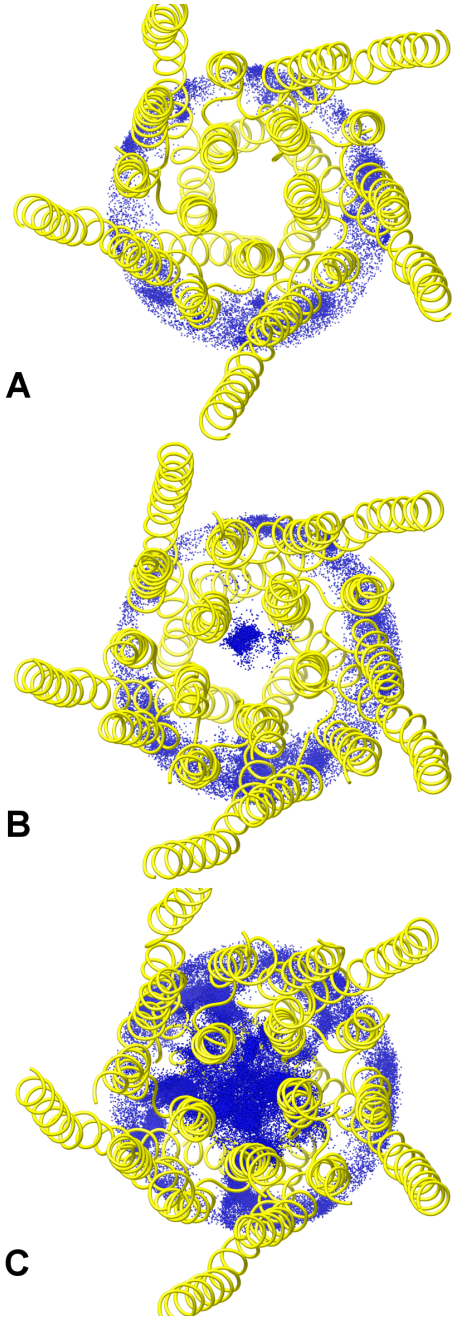


Figure 4.13: **Ion distribution at the ICD level**. Distribution of Na^+ ion positions (blue dots) extracted from the (A) WT, (B) R420D and (C) QDA simulations inside and outside the ICD vestibule (within a cylinder of 20 Å radius, taking the centroid of V264 C_α atoms as the center). The protein average structures are depicted as yellow ribbons.

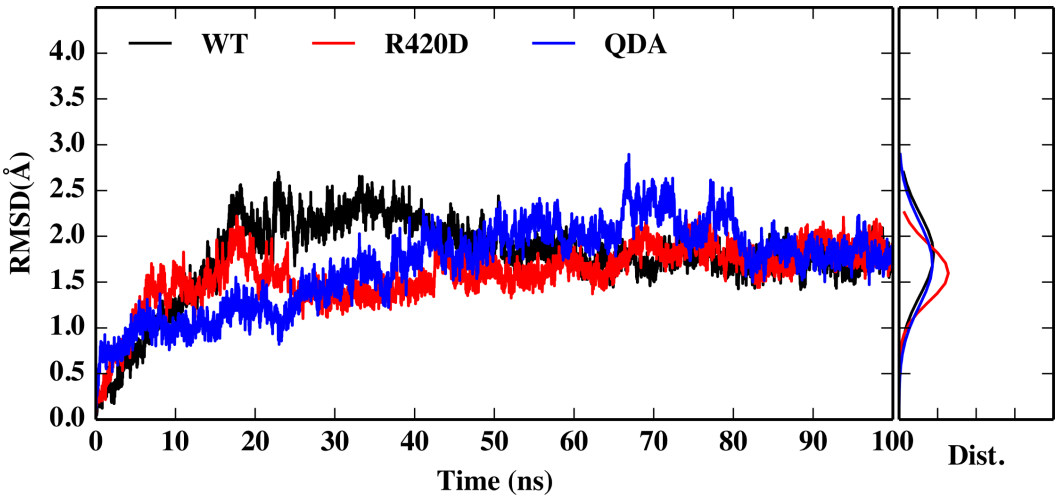


Figure 4.14: RMSD of ICD Portals. RMSD, with respect to the starting configuration, of the stretch of residues that line the ICD portals (302-308, 409-418) in WT, R420D and QDA systems. In the right subplot is shown the corresponding RMSD distribution.

Similarly, the RMSF has not evidenced significant differences in structural fluctuations (Figure 4.15).

Taken together, these results rule against the hypothesis that arginine replacement may impinge on the MA helix dynamics, suggesting, on the contrary, that more localized effects do account for the observed ion permeation. Consequently, we have examined if changes in local electrostatics and/or side-chain structural rearrangements could be the determinants for portal opening. Figure 4.16 shows how the sequential mutation of arginines with smaller residues, either negatively charged or neutral, has induced an increasing portal expansion along with a distinct change in the electrostatic potential of the cavities from mostly positive to mostly negative.

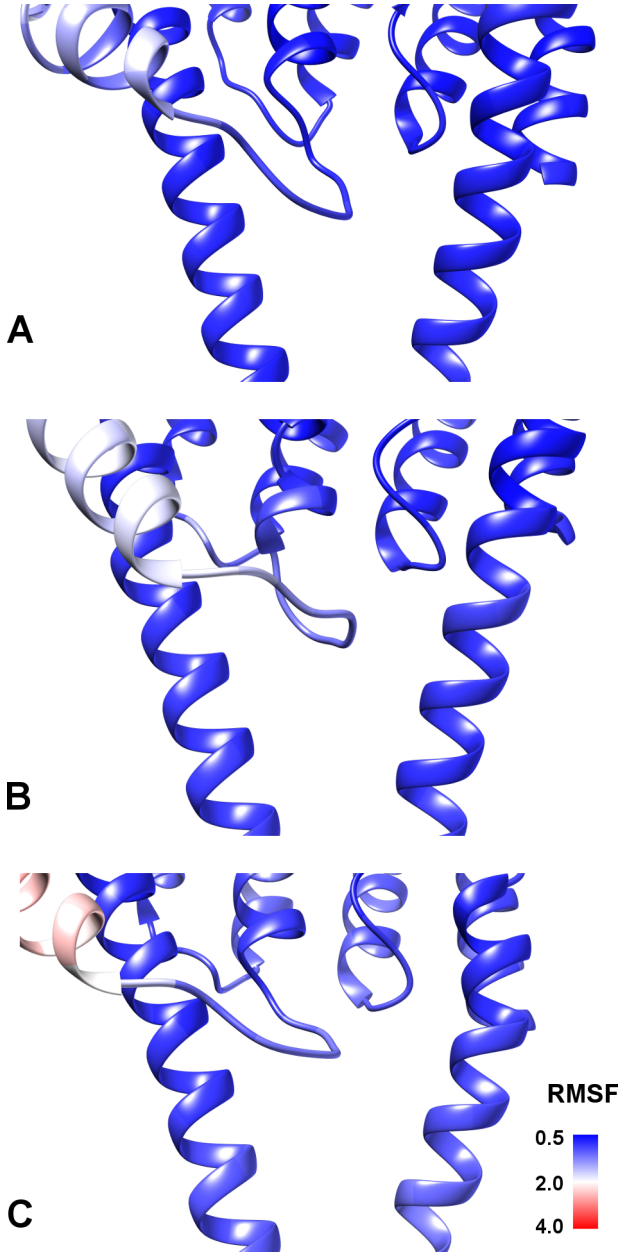


Figure 4.15: Backbone flexibility of ICD portals. Backbone RMSF of the ICD portals in (A) WT, (B) R420D and (C) QDA systems. The protein structure is depicted as cartoons and colored according to the RMSF values.

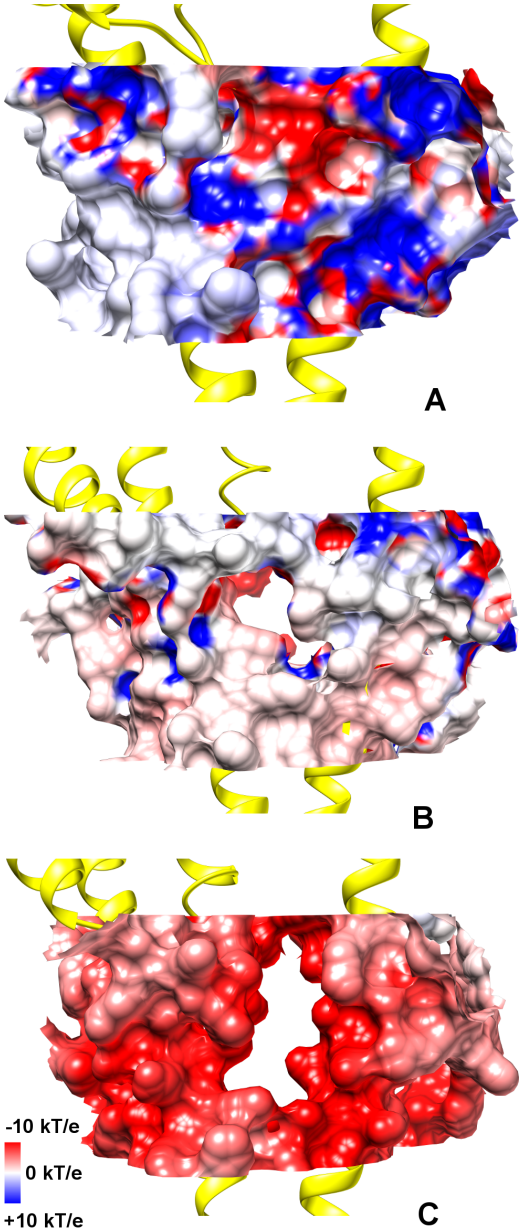


Figure 4.16: **ICD portals.** Electrostatic and steric properties of the ICD portals as issuing from (A) WT, (B) R420D and (C) QDA systems. The ICD portals are depicted as molecular surfaces, colored according to their electrostatic potential from blue (positive) to red (negative). The average structures of M3 helix, post-M3 loop, MX and MA helix, from one subunit, and MA helix from an adjacent subunit are shown in yellow cartoons.

In WT, residue D312 from the post-M3 loop can strongly interact with two arginines of the MA helix, thus anchoring the positively charged side chains of the latter and effectively occluding the lateral portals due to steric hindrance (Figure 4.16A). On the other hand, in R420D mutant, the observed partial expansion (Figure 4.16B) is sufficient to allow ion permeation, though close interactions of residue D312 and D420 with the two remaining arginines (i.e., R416 and R424) are still apparent. In QDA, no more stable electrostatic interactions involving residue D312 are possible and, as a consequence, the mutated residues do not concur to obstruct the ICD portals (Figure 4.16C). Accordingly, going from WT to R420D and QDA, we have observed a gradual increase in the D312 side-chain flexibility (Figure 4.17), which, in turn, reflects the opening of the portals for the reasons sketched above.

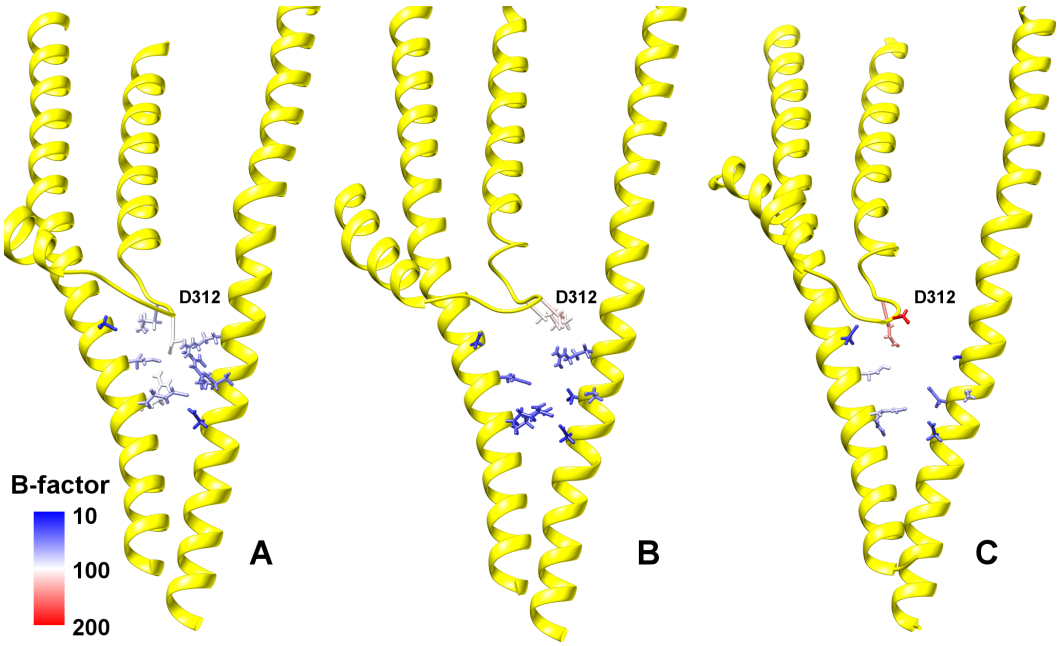


Figure 4.17: Side chain flexibility at ICD portals. View of ICD portals in (A) WT, (B) R420D and (C) QDA systems. Protein is shown as yellow cartoon, and the stretch of residues shaping the portal walls are shown as sticks and colored according to their side chains B-factors. A shift toward red can be appreciated for residue side chains in the post-M3 loop, especially D312.

4.3 Discussion

Atomistic MD simulations of the 5-HT_3A receptor have provided several insights into its structural features and interaction with permeating ions, suggesting interesting connections to channel properties. Although the investigation of a closed-state structure prevents the full comprehension of the role played by some key residues in the conductive conformation, our study, similarly to many other studies on pLGICs, contributed to gain further valuable information on the properties of this channel. Besides, as discussed in the following, results are generally in good agreement with recent experimental findings on the same 5-HT_3A channel and with previous computational studies on homologue proteins.

Our investigation identified a few relatively flexible regions of the protein. In particular, the observed high mobility of loop C is considered crucial for receptor function. It is believed that the outward and inward movement of loop C is pivotal to accommodate the neurotransmitter and to activate the receptor,[209] thus characterizing the protein functional state. Indeed, our simulation showed that in one subunit loop C adopted, for most of the time, an "open-lid" conformation, whereas several partial opening and closing events occurred in loop C of other subunits, in agreement with previous studies.[145,147,204,209-213] Loop F was also found to be flexible as compared to the rest of the protein, especially in the region spanning residues S176 to E186. This loop has been proposed to act as a flexible structure, adopting different conformations suitable for the interactions with ligands of different size.[179] In many crystallization studies of AchBP, this loop was only poorly resolved owing to its disordered structure. Another observed flexible part of the receptor is the M2-M3 loop, which in fact is poorly resolved in the 5-HT_3AR crystal structure. It is worth noting that a number of studies indicated that this loop, along with the interacting region between ECD and TMD, is involved in the complex cascade of movements leading to channel gating.[139,141,152,155,159]

Remarkably, we found that the ECD-TMD interface shapes five inter-subunit lateral cavities, as formed by Cys-Loop, β1- β2 loop, loop F C-terminal part, β10-M1 loop and M2-M3 loop, through which ions can reach the interior of the channel from the extracellular medium. This observation prompted us to re-examine other 5-HT_3AR homologue structures. Surprisingly, we detected similar

fenestrations also in other members of the superfamily: in particular, these structural features are shared by all tested eukaryotic receptors. Since loop F shows the lowest degree of conservation amongst the six binding site loops, it is tempting to speculate that it could play a role in regulating the cavity size in different receptors, depending on its specific residue sequence. In GluCl, the C-terminal part of this loop is folded to form an α-helix turn, splitting the lateral channel into two channels, both narrower than that of nAchR and 5-HT_3A. However, a molecular comprehension on the differences among different receptors and the possible effect of mutations on such lateral pathways demands further investigations.

Overall, the pore dimension along the channel longitudinal axis, as issuing from our MD simulation, was in good agreement with the one estimated from the 5-HT_3A crystal structure:[48] the only two significant differences (K108, E250) have been ascribed to biased side chain conformations of the latter structure. In our simulation, we observed a pore diameter of ~7.5 Å at the most constricted point of the ECD, corresponding to residue K108 (Figure 4.5). Note that this evidence is consistent with the observed smaller permeability of the ECD top side with respect to the lateral channels at the ECD-TMD interface, though the absence of electric field, the relatively limited sampling and the use of a closed structure do not allow to draw a clear distinction between such entrance pathways. Besides, it is known that negatively charged MTS reagents can pass through this region and reach the TMD.[181,214] In any case, our results suggest that residue K108, if mutated, might have a detectable influence on ion translocation, perhaps more than D105, which was previously shown to affect the channel conductance.[215] As expected, the maximum constriction point of the channel was found in the TMD hydrophobic region. Here, the channel size was not shown to change significantly during the simulation, thus forming a pore unsuitable to the passage of solvated ions, in line with the hydrophobic gating hypothesis proposed in homologue proteins.[212]

Moreover, we investigated the complex nature of the cation-selectivity in this receptor. Several studies suggested that the selectivity filter is not located in the ECD, though mutations in this domain were proved to affect single-channel conductance.[140,215] On the other hand, there is a general consensus about the notion that the filter is located at TMD level.[184,216] Our free energy calculations on ion translocation suggested that the filter could be “delocalized” between the extracellular (D271) and the intracellular (E250, D247) side of the

TMD. However, in α7 nAchR, only homologue residues to E250 and D247 were shown to be important for ion selectivity[163,164] and, in 5-HT_3A, the contribution of E250 was shown to be dominant.[184,217] Such an apparent discrepancy between theory and experiments can be reconciled by noting that the PMF analysis was performed on a closed-state channel. Moreover, the most agreed hypothesis on the gating mechanism predicts a peculiar rearrangement of the α-helices shaping the channel. On this regard, according to the known GLIC and GluCl open-state channel structures[202,218] and to the gating model recently proposed in ref.[144], the gating mechanism does involve the tilting of the M2 helices with respect to the channel axis. In particular, channel opening is characterized by an outward polar tilting of the M2 helix upper segment (above residue 264), with the lower segment mostly unaffected. These conformational changes lead to an increase of D271 inter-residue distances, while those among E250 residues remain substantially unaltered. In this scenario, the funnel located at the extracellular side of the TMD gets broadened, thus weakening the possible electrostatic interactions between D271 and the translocating ions. Moreover, the funnel widening allows, in principle, more Na^+ ions to be accommodated, with the result of a more effective screening of D271 residues negative charges. Altogether these two effects may account for a minor contribution of D271 to ion selectivity. On the contrary, the lower portion of the M2 helix (below residue 264) does form a narrow pore, which is essentially similar in both closed and open states. Hence, in the open state residue E250 becomes the paramount determinant for ion selectivity, as overall suggested by experiments on this protein and homologue receptors. It should be noted, however, that the role of this residue on the ion selectivity of all Cys-Loop receptors is still a matter of debate due to contrasting mutagenesis experiments on $GABA_A$ and GlyR.[165,166]

The PMF analysis suggests another interesting feature of this channel. Our PMF profile supported a considerable contribution of residue D247 to ion translocation, along with the nearby residue E250. Therefore, it would be interesting to test if D247 could affect single-channel conductance as well as residue E250, or even ion selectivity, especially in the case of the 5-HT_3AB heteromeric receptor, which is characterized by fewer negative residues at position 250 than the corresponding homomeric receptor. Note that the corresponding residue in nAchR was shown to be important for single-channel conductance[140] (however, since in human 5-HT_3A D247 is replaced by an asparagine, these contributions may not be shared by all organisms). Moreover,

the Na^+ ion attraction by residue D247, while passing through the narrow hydrophobic channel towards the intracellular vestibule, provided an additional driving force to exit the hydrophobic region and the E250 ring. In turn, this may suggest that a multi-ion transport mechanism is not required for ion translocation through the 5-HT_3 channel, as observed in other channels.[219] Nonetheless, a rather contrasting scenario is expected with divalent cations, considering that position 250 was shown to be a divalent ion binding site with a resulting decrease of ionic conductance in both nAchR and 5-HT_3.[217,220] Concerning the obtained trend of the single-ion PMF within the TMD pore, we observe that our results are qualitatively in good agreement with previous calculations on homologue channels[146] (taking into considerations possible differences in the simulation protocols), and possibly more accurate in describing the ending region of the pore owing to a better description of the initial ICD region, as pointed out in the Results section.

Finally, we studied the effect of replacing the MA helix arginine triplet, which is typical of the 5-HT_3A subunit, on cation permeation. Our simulations showed that an increased ionic passage through the ICD portals is achieved by progressively replacing the arginine residues with corresponding residues from the 5-HT_3B subunit. Such a trend nicely correlates, at least qualitatively, with the enhanced single-channel conductance observed while going from the wild-type to the R420D and QDA mutants and further supports the special role of the arginine triplet in determining the characteristic low-conductivity in 5-HT_3A channels, as compared to other homologues.[185] Moreover, our study set out to demonstrate that the hampering effect on ion translocation by the MA helix arginines is due to a combination of electrostatic and steric effects, via the formation of stable side-chain interactions, a result well consistent with an hypothesis based upon cysteine substitution and subsequent modification by MTS reagents.[221] Supported by our findings, we propose that D312 in the post-M3 loop, so far never investigated, plays an important role in stabilizing the arginine triplet in a conformation preventing the cavity opening. However, the present observation is at variance with a recent hypothesis, based on constrained geometric simulations and mutations of MA negatively charged residues, that considers the increase in MA helix flexibility, not observed in our study, as the main cause of the increase of ion conductance.[222] Please note that the increase in conductance reported in the study mentioned above does not necessarily rule out the importance of the electrostatics of the portals. Indeed, two of the three mutated residues, i.e. E434

and D441, being located on different position on the MA helix, actually point towards the exterior of the protein and are solvent exposed. Thereby, their contribution to the conductance could be somehow indirect.

Interestingly, no ion passed through the ICD portals in the WT system within the simulated time interval, though the trypsin-treated 5-HT_3A receptor generated to obtain the crystal structure displayed a higher conductance than the unproteolized receptor. This intriguingly suggests also the possible transmission of (a certain degree of) structural changes from the TMD to the ICD upon gating, as previously conjectured.[223] While a quantitative relationship between the experimental conductance of the trypsin-treated receptor and the extent of portal opening observed in our simulations could not be established, we can reasonably assume that the ICD portal structure, especially the α-helices, undergoes only a slight rearrangement upon gating, as previous experimental studies seem to suggest.[182,206,221,222] If this is the case, the structural differences among mutants observed in this study could be considered relevant also in the channel conductive state.

4.4 Conclusions

Understanding how membrane channels tune ionic flux, regulating the amount and the charge of translocating ions, is pivotal to understand the molecular basis of cellular communication. The 5-HT_3A ionotropic receptor is one of such channels expressed at synapse level in several tissues and, given its involvement in many physiological functions and pathological states, is of systemic importance for the entire organism. Moreover, due to the capability of the 5-HT_3A subunit to assemble into fully functional homomeric channels suitable for *in-vitro* investigation, the 5-HT_3A receptor has emerged as one of the most studied members of the pLGIC superfamily, from which knowledge about the working mechanism of all the other receptors has been obtained.

In addition to the previous knowledge about the structural and biophysical properties of such receptor, the present computational study has revealed a few interesting details concerning the 5-HT_3AR dynamics as well as the ion translocation mechanism and energetics. In particular, this study has unravelled previously unpredicted structural features of this receptors, such as the existence of alternative intersubunit pathways for ion translocation at the interface between

the ECD and the TMD, in addition to the one along the channel main axis. Interestingly, such lateral pathways seem to be also shared by few other pLGICs, suggesting a possible common function for these channels. Moreover this study has provided a molecular interpretation, based on electrostatics and steric hindrance, of the role played by an arginine triplet located in the ICD on determining the characteristic low single channel conductance of the 5-HT_3A receptor as compared to other members of the superfamily. Finally, the molecular dynamics simulations and the PMF analysis altogether suggest a possible impingement of few residues lining the pore interior, so far never investigated, on the ionic translocation process along the channel, in each of the three domains of the receptor. This may ultimately also impinge on the conductive properties of the receptor, particularly its cation vs anion selectivity and channel conductance, pinpointing them as valuable targets for further mutagenesis investigations.

Chapter 5

The MscL channel

Mechanosensitive channels are integral proteins located on the bacterial cytoplasmic membrane, whose function is that of a natural pressure valve preventing cell lysis in case of hypoosmotic stress.[224,225] Among the members of this protein family, the mechanosensitive channel of large conductance from *E.coli* (Eco-MscL) is the best-known protein channel.[32,50] In response to increased membrane tension, Eco-MscL can form a large pore with a diameter of up to ~30 Å. In the open state, the MscL pore has a non-selective conductance of 3 nS and it allows the passage of large hydrated solutes and small proteins down the concentration gradient. The structure of a MscL homologue from Mycobacterium tuberculosis (Tb-MscL), resolved by X-ray crystallography, has shown a symmetric homo-pentamer,[42,226] in which each subunit consists of a short N-terminal helix, two membrane spanning helices (named TM1 and TM2) connected by a periplasmic loop and a cytoplasmic helix (Figure 5.1A). The five TM1 helices form the pore lumen that is characterized by a hydrophobic stretch of residues (I14-V21) creating a constriction at the cytoplasmic side (Figure 5.1B), while the TM2 helices form contacts with the surrounding lipid bilayer. The X-ray structure of Tb-MscL represents the closed-state configuration of the protein channel with a radius of ~2 Å along the narrowest region of the pore and provides the framework with which extensive mutagenic and functional data have been interpreted.

Mutagenesis experiments in Eco-MscL identified residues L19-V23 to be critical for channel gating;[227] the corresponding residues (L17-V21) in Tb-MscL are located in the narrowest part of the pore (Figure 5.1B). The selective

mutation of these residues increased the feasibility of channel opening, leading to slow/no-growth gain of function phenotypes.[228] Further, systematic substitutions of G22 in Eco-MscL by all other amino acids showed that hydrophilic residues at this position resulted in channel gating at reduced tension with respect to the wild-type, while hydrophobic residues made the channel harder to open.[229] Several studies showed that the introduction of charged residues at the same position led to an activated channel in the absence of any applied tension (e.g., mutating the hydrophobic pore residues to cysteine and attaching a charged MTS reagents.[230-232] These findings, together with the ability of MscL to form a large pore and assemble into a fully functional protein when reconstituted into synthetic lipids, led to conceive MscL as an ideal candidate to function as a triggered nanovalve. Indeed, recently engineered MscL channels have been designed to respond to stimuli like pH or light in applications tailored to molecular sensing or drug delivery.[77,233,234] The light controlled channel gating is particularly interesting as it provides a finer temporal control and does not require any change in the system environment, as reported recently by Koçer *et al.* [77]: in this case, a light-triggered gating has been artificially introduced by a photo-sensitive functional group alkylated with a purposely mutated cysteine in the pore interior (Figure 5.1C). Hence, it has been shown how the application of light has induced a spontaneous channel opening, in the absence of membrane tension, according to a designed charge-repulsion based trigger. On the other hand, in more recent studies by Koçer and coworkers[235,236] it has been shown that MscL channel gating can be also triggered by just one charged group in a single subunit, therefore following a charge-induced hydrophobic breaking mechanism. Overall, these experiments have demonstrated the possibility to I) easily manipulate MscL pore size, II) control the flux of charged compounds and III) deliver bioactive molecules into the cell, thus showing the versatile use of MscL as a controllable nanoscopic valve.

In light of such remarkable applications, several aspects of the charge-induced gating mechanism deserve further investigation, among which I) the scale of protein conformational changes upon charge introduction, II) the extent of the channel structural expansion and III) the selectivity and geometric limitations of molecular transport across the channel. Such details demand an atomistic level understanding of the interplay between the induced electrostatic effects and the protein structural rearrangements. In this regard, MD simulations can be profitably used to provide structural characteristics and dynamical features of

protein channels, while retaining explicitly the environmental effects mainly due to the lipid bilayer and the solvent. Previous computational studies have focused on the gating conformation transitions of the wild-type MscL protein [237-240], the mutation effects in the pore hydrophobic region [241] and the lipid composition effects [242,243]. In the present study, we have performed atomistic MD simulations to investigate the effects of charge incorporation on the hydrophobic pore of Tb-MscL. In analogy with the experimental setup of ref. [77], five photosensitive ligands, one for each protein chain, have been incorporated into a constricted site of a Tb-MscL mutant (V21C). Then, each ligand was sequentially replaced by a charged acetate group (Figure 5.1D) and the resulting systems simulated independently.

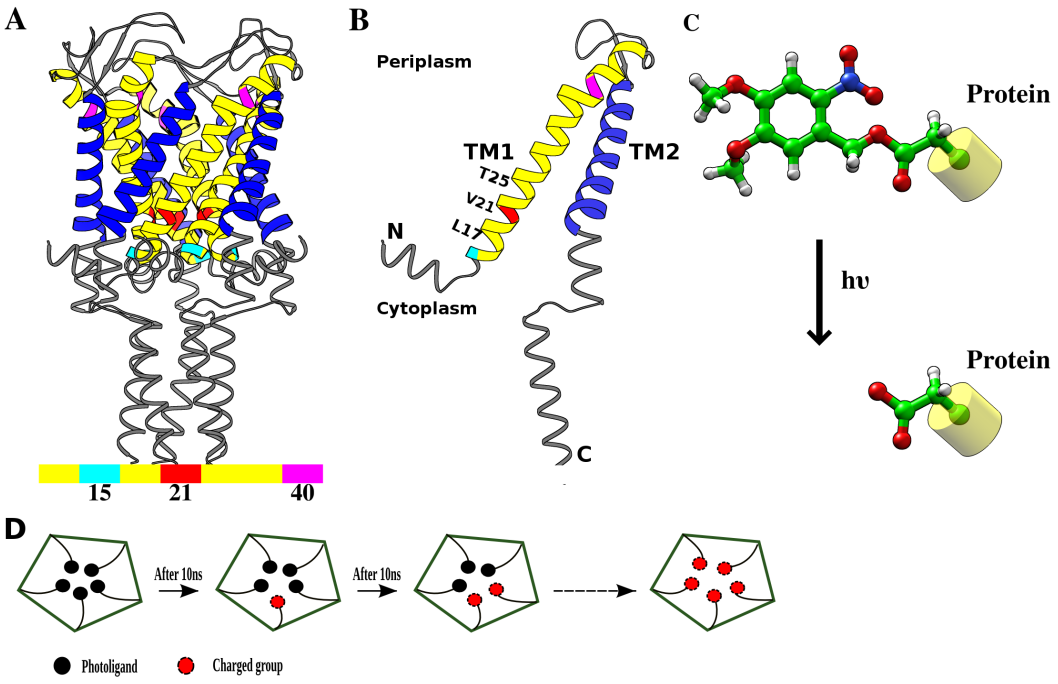


Figure 5.1: Structure of Tb-MscL and photo-sensitive compound. (A) X-ray crystallographic structure of Tb-MscL (PDB ID: 2OAR) with the transmembrane helices TM1 and TM2 colored in yellow and blue, respectively. Locations of residues lining the inner helix (TM1) are indicated. (B) Single subunit of Tb-MscL. (C) Structure of photo-sensitive compound that releases the charged group upon light irradiation. (D) Schematic view of the step-by-step substitution process followed in the MD simulation protocol.

The structural features of the engineered models have been compared with a wild-type model of Tb-MscL. Further, the effects of charge incorporation at two alternative sites has also been examined, considering the L17C and T25C mutants. In addition, we considered a different Tb-MscL model, in which a charged group was attached to only one subunit, mimicking the charge-induced hydrophobic breaking process previously reported.[235,236] The obtained results report new molecular details on channel structural opening, ion permeation properties, asymmetric subunit motions, and site sensitivity towards charge perturbations, therefore providing further insights into MscL charge-induced gating.

5.1 Computational details

5.1.1 Systems preparation

In ref. [77], light-actuation of MscL has been achieved attaching a photo-sensitive compound (6-nitroveratryl alcohol), which splits into 6-nitrosoveratyl aldehyde and a free acid upon light irradiation, to a G22C mutant of Eco-MscL through a cysteine-selective alkylating reagent. UV photolysis resulted in cysteine-bound acetates (hereafter referred to as the charged group) in the pore lumen (Figure 5.1C) and consequent channel gating. Here, we attached the same photo-active ligand in all subunits at residue 21, the most constricted site of the Tb-MscL. The starting structure of the MscL channel was embedded into a homogeneous lipid bilayer of DOPC (1,2-dioleoyl-*sn*-glycero-3-phosphocholine) spanning ~100 Å along the lateral dimensions (x,y) using the CHARMM-GUI server,[187] followed by solvation with TIP3P water, extending up to 30 Å from the solute in the axial dimension (z-axis) using VMD.[188] Since the constricted site is extremely narrow to fit the ligands without steric conflicts, the channel was slowly expanded through multiple short simulations by placing an uncharged LJ atom at residue C21 centroid and increasing its LJ radius parameter. Once sufficient space was created, the ligands were attached to C21 in all five subunits. K^+ and Cl^- were added to the system, setting the ion concentration to about 1 M. The engineered model with five ligands (hereafter referred to as 5L) was then minimized to remove steric clashes, slowly heated up from 100 K to 300 K in

about 1 ns simulation and then equilibrated for 10 ns in a NPT ensemble. During equilibration, the pressure was kept at 1 atm using the Langevin piston method with a piston period, damping coefficient and piston temperature of 100fs, 50fs and 300K, respectively. Different engineered models were then sequentially generated by replacing each ligand by the charged group (Figure 5.1D). Before each replacement step (every 10 ns of the production phase), the starting models were equilibrated for 6-8ns. To verify the charge effects on the pore, a model representing the wild-type (WT) protein was simulated using a snapshot of NL model (obtained after 50 ns) as the starting configuration and replacing the charged group in all subunits with valine. To investigate the charge activation of a single subunit in a different MscL model, we extracted an equilibrated configuration of the WT model and mutated residue V21 to cysteine in a single subunit, to which the same charged group was attached. Other simulations involving alternative sites were initiated from the same configuration of the WT model, after mutating the corresponding residues to cysteine (namely, L17C and T25C) and attaching the corresponding charged group in all subunits.

5.1.2 Simulation and analysis details

Molecular Dynamics. A summary of the simulated systems is reported in Table 1. All production runs were performed in a NVT ensemble, applying a homogeneous external electric field (E_z), proportional to a defined voltage (V = 1 Volt; $E_z = -V/L_z$), along the z-axis (L_z) perpendicular to the membrane. The simulations were performed enforcing PBC using NAMD (v2.9) [190] and the CHARMM force field for the protein (v27) and lipids (v36).[244-246] The ligand atomic charges were obtained via fitting with a quantum-mechanically derived electrostatic potential issuing from B3LYP/6-31G* calculations using the Gaussian package,[247] according to the restrained electrostatic potential (RESP) procedure using the RED software tools [248]. The parameters for the ligand and the acetate groups were adopted from the CHARMM CGEN small molecule force field. [249] All covalent bonds with hydrogen atoms were kept rigid using the SHAKE algorithm, allowing the use of a 2 fs time step for the integration of the equations of motion. A cutoff radius of 12 Å was used for non-bonded interactions, applying a smoothing function beyond a distance of 10 Ang. The long-range electrostatic interactions were evaluated with the PME [107] method,

whereas a Langevin dynamics was used to maintain the system temperature at 300 K.

Models	Description (charge[a])	Time (ns)
5L	5 Ligands at position 21 (0e)	30
4L	4 Ligands ,, (-1e)	30
3L	3 Ligands ,, (-2e)	30
2L	2 Ligands ,, (-3e)	30
1L	1 Ligand ,, (-4e)	50
NL	No Ligand ,, (-5e)	100
WT	Wild type (0e)	50
WT_1e	WT with charge in a single chain (-1e)	80
17M	5 charges at position 17 (-5e)	50
25M	5 charges at position 25 (-5e)	50

Table 1: Summary and description of the simulations.

[a]Total charge at the functionalized site.

Free energy calculations. A computational approach similar to that reported in previous studies [143,159] was employed to estimate the PMF governing K^+ translocation through the channel using the ABF methodology, as implemented in NAMD.[126,128] The PMF was evaluated along the z-coordinate (channel axis), considering a region that extends up to ~10 Å on either side of the functionalized site. The simulations were performed in 10 separate windows of 2 Å length for about 11 ns. For each window, the starting coordinates were extracted from the production runs described above. Excess ions were removed, retaining only those required for charge neutralization including a single K^+ ion in the pore. The ions in the bulk solution were restrained to obtain the single-ion PMF profile. Within each window the average force acting on the selected ion was accumulated in 0.1 Å sized bins, and the biasing force was applied after 800 samples.

Trajectory analysis. RMSDs were computed performing a least-square fitting to a reference structure, either the starting one or the X-ray crystallographic

structure. The pore radius profile was obtained using the HOLE program.[195] The ion flux was estimated by calculating the number of ions entered from one side and exited through the opposite side of the channel (considering the TM domain only) with an in-house code written using MDAnalysis.[196] The analysis was performed omitting the first 10 ns of the MD trajectory during which the models undergo structural rearrangements. The pore area was approximated by the area of a regular pentagon by setting as its edge length the average distance between C_α atoms in adjacent subunits. Water and ions around a permeating K^+ were estimated from the average number of corresponding species within a predefined cut-off radius, as a function of the z-coordinate. Channel conductance was estimated from the ratio of the average current and applied voltage.[250,251] Inter-subunit contacts were obtained considering the C_α atom distances. Trajectory visualization was performed using the VMD software.[188] Figures and plots were generated using UCSF-Chimera and Matplotlib software.[197,198] Unless stated otherwise, the axial positions in all plots are indicated with respect to the origin fixed at the C_α geometric centroid of residue 21.

5.2 Results

5.2.1 Structural rearrangements of the engineered MscL

The Tb-MscL protein is a relatively small homopentamer channel of 151 residues (17KDa) (Figure 5.1A). In particular, each monomer has two transmembrane α-helices, namely TM1 and TM2, where TM1, one of the most conserved motifs of the mechanosensitive channel family (V^{15}DLAVAVVIGTAFTALVTKFTDSIIT^{40}), forms the inner core of the channel and is mostly hydrophobic. First, we have analyzed and compared the structures of different engineered MscL channels issuing from the sequential replacement of the attached (neutral) photo-sensitive ligands, i.e. 6-nitroveratryl alcohol, by negatively charged acetate groups. It should be noted that in the original experimental work[77] only the MscL channels carrying either the five neutral ligands (5L system) or the corresponding photoproducts (NL system) were presumably detected. Nevertheless, our purpose was to study in some detail the effect of each charge insertion on the MscL channel properties, so as to

investigate the virtual realization of a multi-state gating mechanism (a proof-of-principle of a multi-state MscL channel activation was provided in ref. [235], even if the gating mechanism was not controllable by light). For such a purpose, all systems were subjected to an applied voltage (1 Volt) and a 1 M KCl salt concentration.

The stability of the simulated MscL models was assessed by evaluating the structural deviations from starting configurations. RMSD of the backbone atoms, depicted as function of simulation time, have shown a plateau after about 10 ns (Figure 5.2), following an initial increase due to structural relaxation. The distribution of RMSD values over the last 20 ns has shown an average deviation of ~2 Å (Figure 5.2 and Table 2). The RMSD from the X-ray structure of the whole protein were generally higher (about ~3 Å), but considerably reduced when the TM or C-terminal domains were separately examined (Table 2). In summary, all models displayed an apparent convergence towards an equilibrium conformation in a few nanoseconds and a stable structural arrangement afterwards during the simulations.

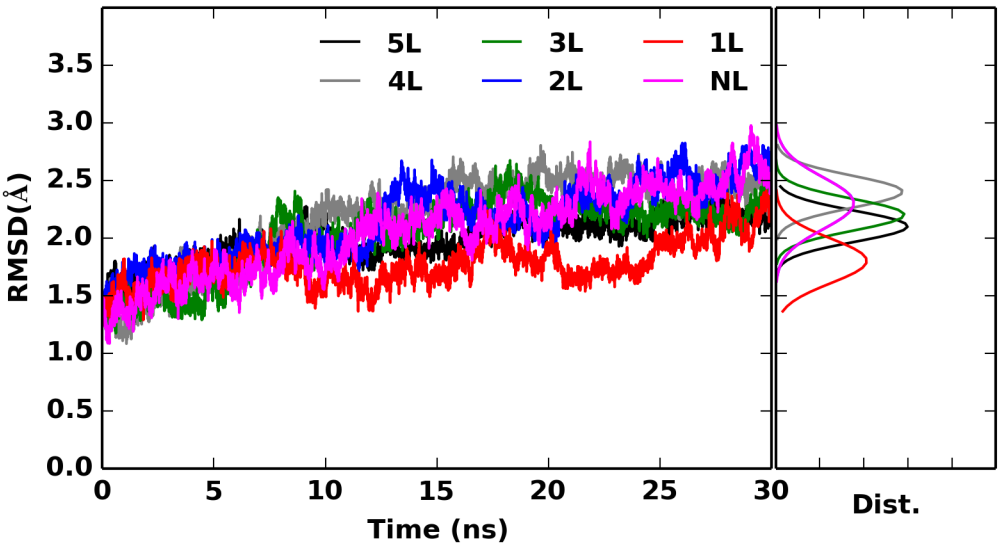


Figure 5.2: RMSD as a function of time. Backbone RMSD, calculated with respect to the starting configuration of respective simulations. The right panel represents the distribution of the RMSD values over the last 20 ns.

Model	A	B	C	D
5L	2.1 (0.13)	2.4 (0.08)	1.6 (0.10)	1.0 (0.11)
4L	2.4 (0.14)	2.9 (0.12)	1.7 (0.08)	0.9 (0.11)
3L	2.2 (0.14)	2.9 (0.12)	1.8 (0.10)	0.9 (0.12)
2L	2.3 (0.22)	3.2 (0.18)	1.8 (0.10)	1.0 (0.17)
1L	1.9 (0.20)	2.9 (0.18)	1.9 (0.08)	0.9 (0.12)
NL	2.3 (0.22)	3.4 (0.17)	1.8 (0.08)	1.5 (0.13)

Table 2: Backbone RMSD (in Å) of the engineered MscL channels. RMSD evaluated with respect to (A) the starting configurations and the X-ray structure for (B) whole protein, (C) transmembrane and (D) C-terminal domains. Standard deviations are in parenthesis.

Moreover, the flexibility of the MscL protein was examined by evaluating the backbone structural fluctuations. To this end, the RMSFs mapped onto the average structures (Figure 5.3) have shown that the most mobile regions are localized mainly on the periplasmic loop of the protein. The smaller fluctuations in the TM region indicated the structural integrity of this region upon ligand or charge incorporation.

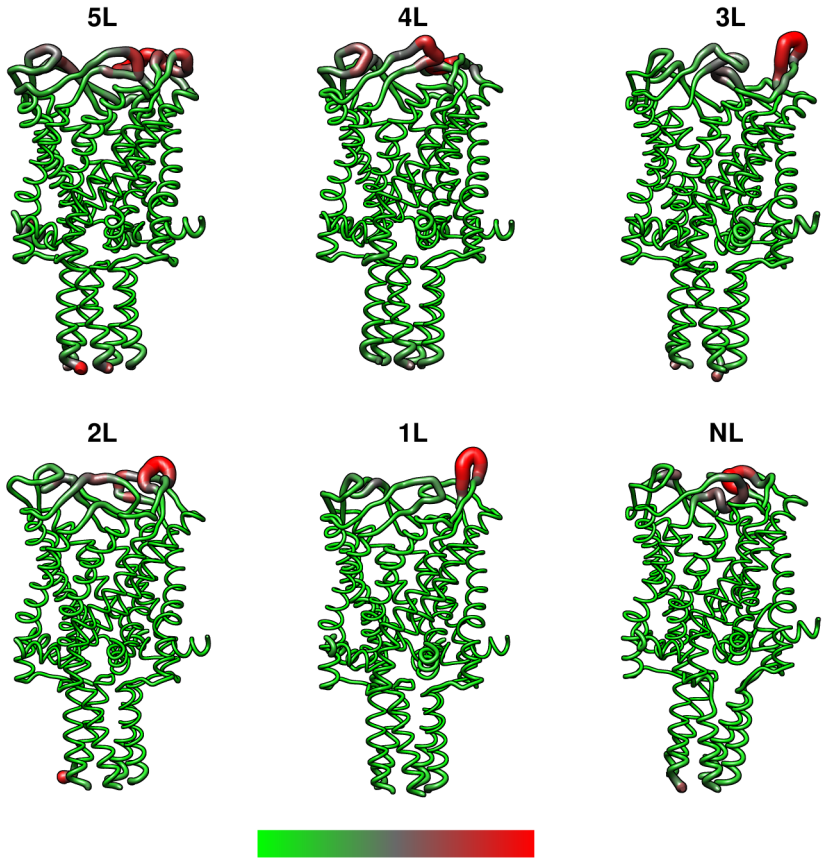


Figure 5.3: Per-residue averaged root mean square fluctuations. Backbone RMSF mapped onto the average structure. The protein is represented as ribbon whose color (green to red) and thickness is proportional to the RMSF values.

5.2.2 Structural expansion of the channel

The extent of channel structural expansion in each engineered MscL model was examined by evaluating the average pore radius along the channel axis, considering only the backbone atoms for the sake of comparison (i.e. the steric effects of side chains and ligands were neglected). In Figure 5.4, the average pore radius as a function of the axial position is depicted for all models, including the wild-type protein.

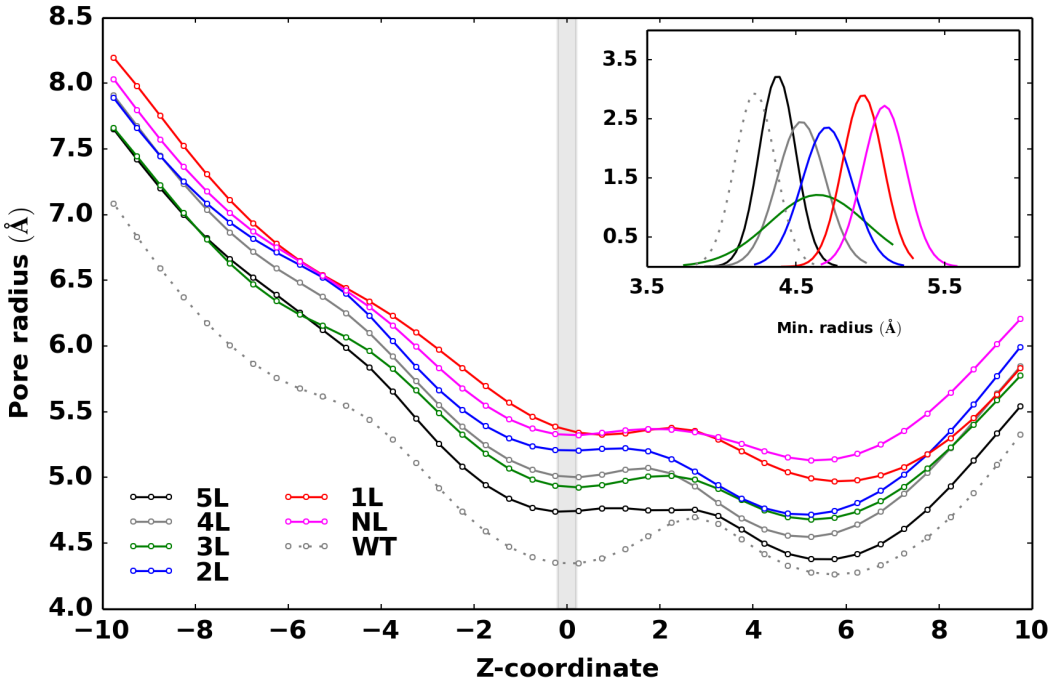


Figure 5.4: Average pore radius along the channel. Average pore radius along the channel axial position (Z-coordinate) with respect to the mutated site (shaded portion, Z = 0 Å). Inset plot represents the distribution of minimum pore radius calculated by limiting the analysis to the constricted zone of the pore (residues 17 to 23).

The pore radius profile has shown a gradual expansion of the channel following the sequential incorporation of additional negative charges into the channel core. A rather similar degree of expansion is observed along the whole TM1 helix, suggesting its tendency to behave as a rigid body. The distribution of the minimum pore radius (see inset of Figure 5.4) has provided a monotonic shift (5L<4L<3L<2L<1L<NL) towards larger radius values in going from 5L to NL. Hence, the observed trend in pore radius supported an increasing charge effect, as previously observed on charge-mediated MscL gating.[235] To further estimate the extent of channel opening, we have approximated the area of the pore by evaluating the area of a regular pentagon with edge given by the average distances of adjacent C_α atoms (i.e., on proximal subunits) at the same residue along a stretch of aminoacids flanking the functionalized site. Results depicted in Figure 5.5, once again indicated a larger expansion of the NL model as compared to the 5L, in line with the pore radius profile (Figure 5.4).

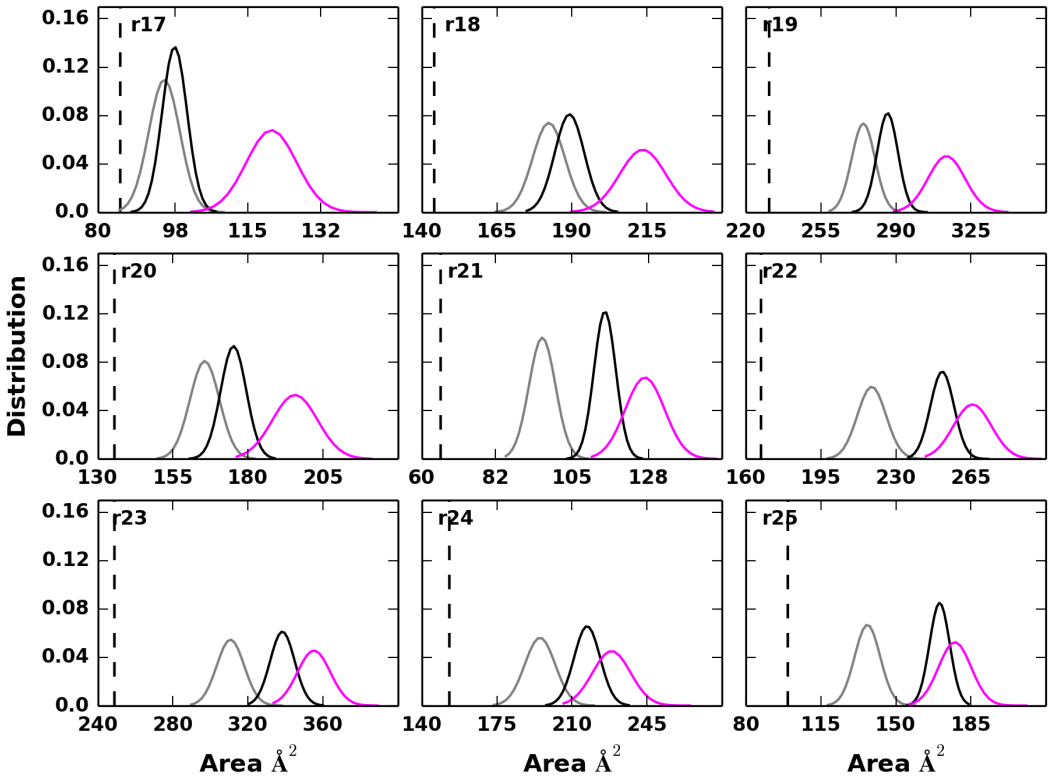


Figure 5.5: Distribution of the pore area per residue. Pore area estimated for a stretch of residues lining the TM1 helix for WT (gray), 5L (black) and NL (magenta). The vertical dashed line corresponds to the X-ray structure.

However, the maximum channel opening was quite smaller than the one reported for membrane tension and charge-induced mechanism (up to ~30 Å diameter).[252,253] In order to test the overall stability and geometric features of the expanded channel in the NL model, an additional 100 ns MD simulation was performed at high temperature (330 K), but no significant changes have been noticed.

Moreover, it is worth nothing that the WT simulation was initiated from an equilibrated configuration of the NL model after replacing each charged group at the functionalized sites with valine. Such a hydrophobic substitution led to a fast channel closure (a few nanoseconds) as indicated by the radius profile (Figure 5.4). A similar spontaneous channel closure has been observed for MscS channel during unrestrained MD simulations.[254] Note that the pore dimension in the

WT model appeared somewhat larger than the corresponding X-ray structure, owing to thermal fluctuations and a more realistic environment (crystal packing forces are generally unleashed in MD simulations of proteins, as observed in a previous MD study on the X-ray structure of Tb-MscL[255]).

5.2.3 Ion permeation: PMF and translocation mechanism

The ion flux through the functionalized MscL proteins was analyzed by monitoring separately the total number of ions entering into the channel (Entry in Table 3) and the number of ions permeating the channel through opposite vestibules of the pore (Exit in Table 3). Then, a permeation rate has been obtained as the ratio of permeated over entered ions. In all cases, results have been collected from corresponding MD simulations upon application of voltage. Due to the overall small pore dimension and the filtering effect of the negatively charged groups, only K^+ ions were able to permeate through the engineered MscL channels from the periplasmic side. As reported in Table 3, the permeation statistics for K^+ ions has shown the occurrence of ion flux only through 1L and NL models. In 1L model, the bulky ligand in the pore largely attenuated the permeation events (~11%), albeit a structural expansion similar to the NL model is observed (Figure 5.4). The NL model allowed the highest ion flux (~57%) throughout the entire simulation (a rather similar permeation rate has been also obtained from selected portion of the whole trajectory). In WT simulation, no ion permeation was observed under applied voltage.

Model	Entry (#)	Exit (#)	Permeation (%)
5L	42	0	0
4L	32	0	0
3L	36	0	0
2L	29	0	0
1L	79	9	11
NL	129	77	60
WT_1e	86	33	38

Table 3: Ion flux (K^+) through the engineered MscL channels

Energetic barrier for K^+ translocation through 1L and NL models that permit the ion flux was examined by evaluating the single-ion PMF profile along the pore, considering a region that flanks ~10 Å on either side of the mutated site. The PMF profile (Figure 5.6A) reports a downhill in the free energy barrier in both systems as the ion approaches the mutated sites from the periplasmic side (Figure 5.6A, shaded region), where it is highly stabilized by the electrostatic attractions of the charged groups. In both cases, an increasing PMF barrier is observed as the ion exited from the cytoplasmic side since this region constitutes a hydrophobic stretch. The higher barrier observed for the 1L model at this stretch can be rationalized as follows: the bulky ligand narrows the channel size and perturbs the local symmetry of the pore lining residues, thereby hindering the linear flux at the exit pathway from the functionalized site (Figure 5.6B), providing an additional energetic cost for translocation. On the other hand, the NL model showed a more attractive PMF (~ 2-3 kcal/mol) in the $z < 0$ region as compared to the 1L model, due to the higher total charge at the functionalized site and the absence of any bulky ligand.

As evidenced by the single-ion PMF, the electrostatic attractions extended by the ring of charged groups may present an electrostatic trap, restraining the K^+ ion movement along the pore. To investigate the factors that facilitate the escape from such a strong attractive site, the ion coordination of permeating K^+ ions has been carefully scrutinized, considering both water and other cations. For each K^+ ion inside the pore, the average number of water/other K^+ ions within a spherical region of defined radius has been evaluated. The number of water molecules around the ion has shown to decrease as it approaches the mutated site (Figure 5.6C) and such dehydration is similar for both the 1L and NL models. It is reasonable to assume that the loss of water is compensated by the interactions with the charged groups, specifically the oxygen atoms. At variance, the average number of surrounding cations has increased around the mutated site (Figure 5.6D). The NL model has shown an average number of up to ~2.5 cations within a sphere of 5 Å radius (~1 cations within 4 Å radius, see inset in Figure 5.6D). The surrounding cations can give a significant contribution to the permeating ion by counterbalancing the attractive forces of the negatively charged ring, thus aiding the ion to escape from the electrostatic trap described above. In 1L model, the bulky ligand restricts the available pore size and consequently the number of surrounding cations (~1 cations within 5 Å radius). This result is also confirmed

by the average number of K^+ ions along the pore (see the histogram of K^+ counts along the axial positions in Figure 5.7).

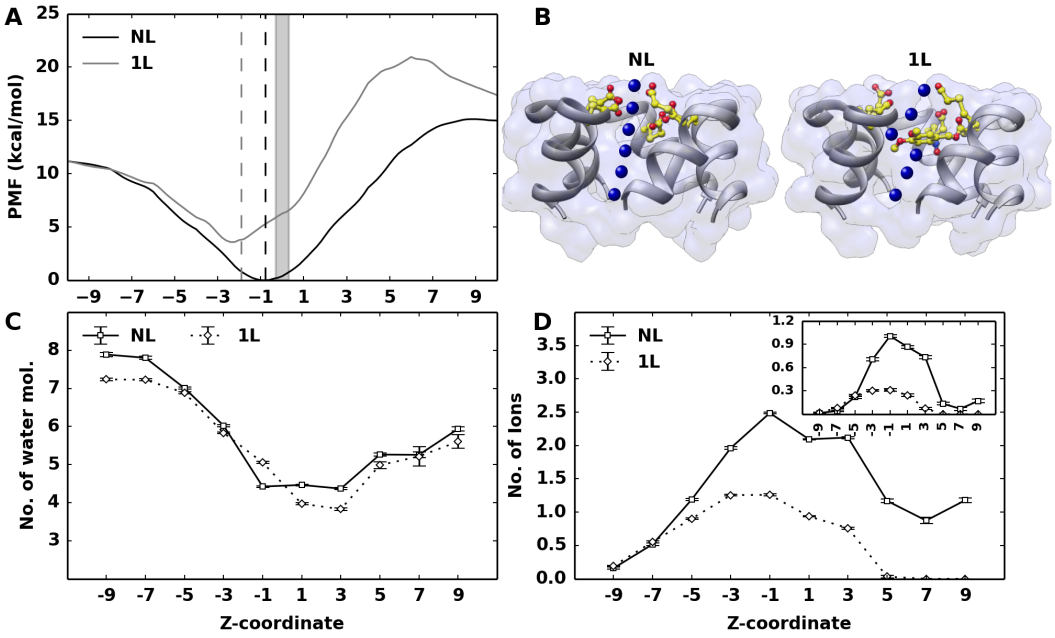


Figure 5.6: PMF profile, water and ion coordination around the potassium ion. (A) PMF profile as a function of the z-coordinate. The origin was set at the C_α geometric centroid of the mutated site. Vertical dashed lines indicate the average location of the acetate group oxygens centroid with respect to the origin. (B) Schematic view of ion exit pathway from the mutated site. The charged group (NL) and the bulky ligand (1L) are shown in yellow. Blue spheres inside the pore represents the K^+ ion. (C) Average number of water molecules and (D) potassium ions within a radius of 3.5 and 5Å, respectively, around the permeating K^+ ion. Data obtained with a smaller radius (4 Å) is shown in the inset.

The lower cation density at the level of the functionalized site in the 1L model is expected to provide a less effective electrostatic counterbalance, as required for a rapid escape from the highly charged site of the pore. Such a result, in turn, may explain the lower permeation rate observed in the 1L model with respect to the NL model (Table 3).

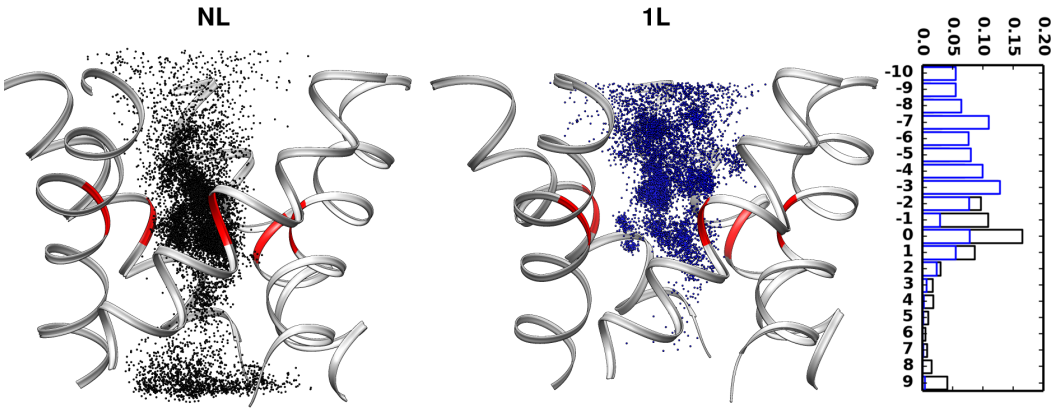


Figure 5.7: Ion occupancy and histogram of ion counts along the axial positions. Occupancy of K^+ ion inside the pore obtained by extracting the ions within a distance of 10Å from the mutated site along the radial and axial dimensions. Right panel shows the histogram of K^+ counts along the axial positions for NL (black) and 1L (blue) models.

5.2.4 Alternative functionalization sites

The TM1 helix in Tb-MscL contains other two hydrophobic residues (L17, T25) similarly exposed to the pore with respect to V21, and located on either side of V21. We have investigated the effect of charge incorporation at these alternative sites (Table 1), performing additional simulations of corresponding fully charged systems (i.e., after introduction of the charged group in all five subunits). To be specific, we have generated two additional MscL protein models by attaching the same functional group, i.e. a charged acetate group, at either position 17 or 25 and compared the results with the original engineered protein (NL system). The pore radius profile along the channel axis, as issuing from both models, has shown no appreciable expansion as compared to the result obtained for position 21 (Figure 5.8).

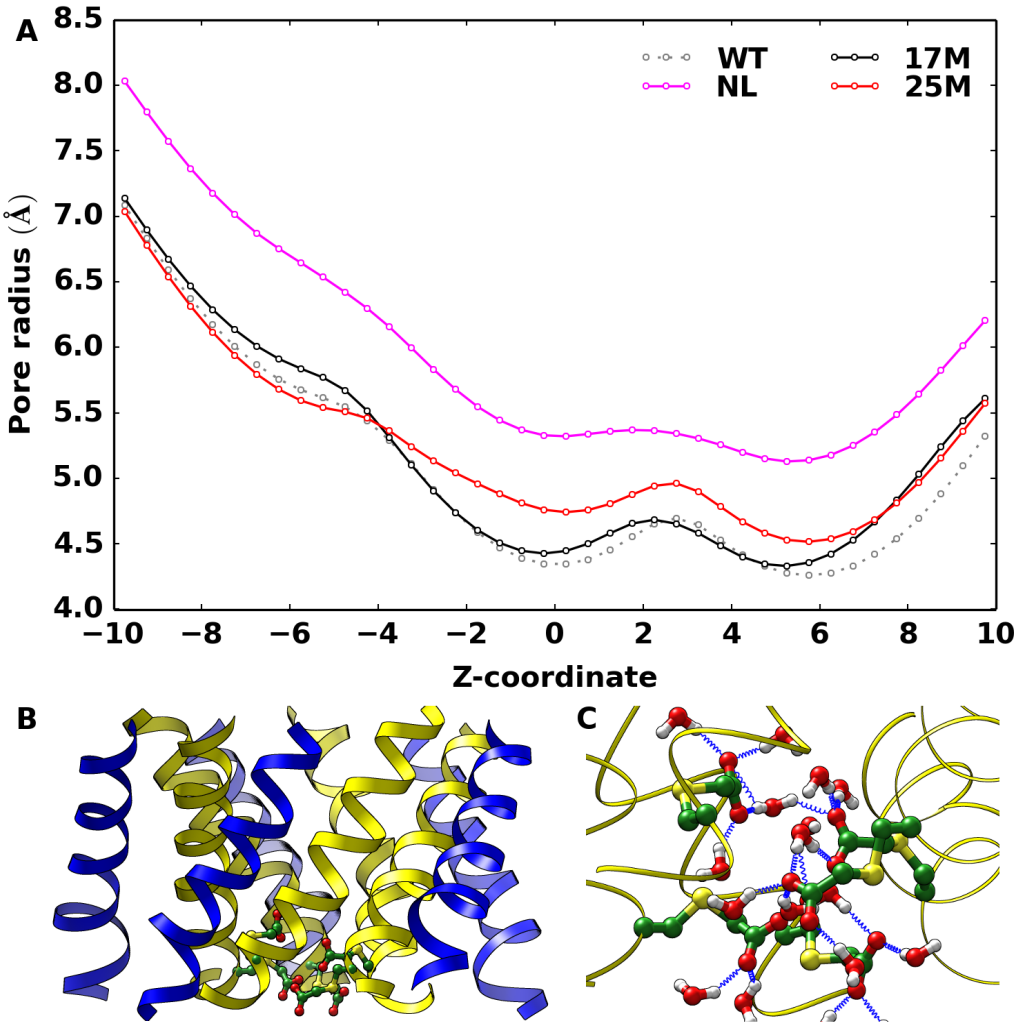


Figure 5.8: Average pore radius and protein-water interaction at cytoplasmic entrance. (A) Average pore radius along the axial position. (B) Representative snapshot of the 17M model with charged groups shown in green. (C) Interaction of charged groups with water molecules within a distance of 3 Å (17M model). Protein-water hydrogen bonds are shown as blue springs.

Overall, the tested models displayed a pore radius similar to the WT protein, indicating the absence of any significant structural transitions induced by the charges. Note that the location of T25 already provides sufficient space to accommodate the charged groups without demanding for any additional structural rearrangements. As a matter of fact, the same pore size has been observed at

position 21 and 25 in both corresponding engineered proteins, namely NL and 25M models, with the difference that in the former case a channel expansion has occurred that allows the coordination of K^+ ions at the pore center. On contrary, residue L17, being located at the narrower entrance at the cytoplasmic edge is expected to be sensitive to charge perturbations (Figure 5.8B). However, visual inspection revealed that the charged groups inserted at this position reorient themselves to make stronger interactions with the water molecules that fill the space between the transmembrane and the cytoplasmic domain (Figure 5.8C). The interaction with water molecules can effectively shield the charge repulsions, thereby attenuating the need for an energetically expensive structural change. As a result of such structural differences, the ionic transport capability was drastically reduced in both alternative functionalized MscL proteins, leading to about 5-7 permeated K^+ ions within a 40 ns time interval, against 42 K^+ ions for the NL system within the same time interval.

5.2.5 Single-subunit engineered channel

Recent studies have demonstrated that introducing only one charge in a single subunit of the MscL hydrophobic pore (at position G22 in Eco-MscL) is sufficient to trigger channel gating in the absence of membrane tension.[235,236] Note that this is a rather different system with respect to 4L, which is also characterized by the presence of one charge in the channel core region, because in the latter case four bulky ligands prevent channel opening. Such a gating mechanism is driven by a rather peculiar phenomenon, in contrast to the charge repulsion effect described above. In fact, it has been suggested that the incorporation of one charge in the channel interior does break the short-range hydrophobic interactions by forcing hydration inside the pore.[256,257] It is reasonable to assume that already a few water molecules approaching the charged site may be enough to significantly weaken the inter-subunit interactions at the level of the most constricted channel region. Besides, experiments suggested that the partially open structure achieved via single-subunit charging is the result of an asymmetric movement of the modified subunit with respect to the others.[235,236]

In order to better investigate, at molecular level, the present gating mechanism, we conceived an additional model (WT_1e) in which the same negatively charged group was directly attached to the constricted site (V21) in a

single subunit (subunit 1) of the WT protein, therefore mimicking the light-actuation of a single subunit of MscL. Inspection of the average pore radius along the channel axis revealed a clear expansion of the WT_1e model with respect to the WT protein (Figure 5.9), albeit to a lesser extent compared to the NL model.

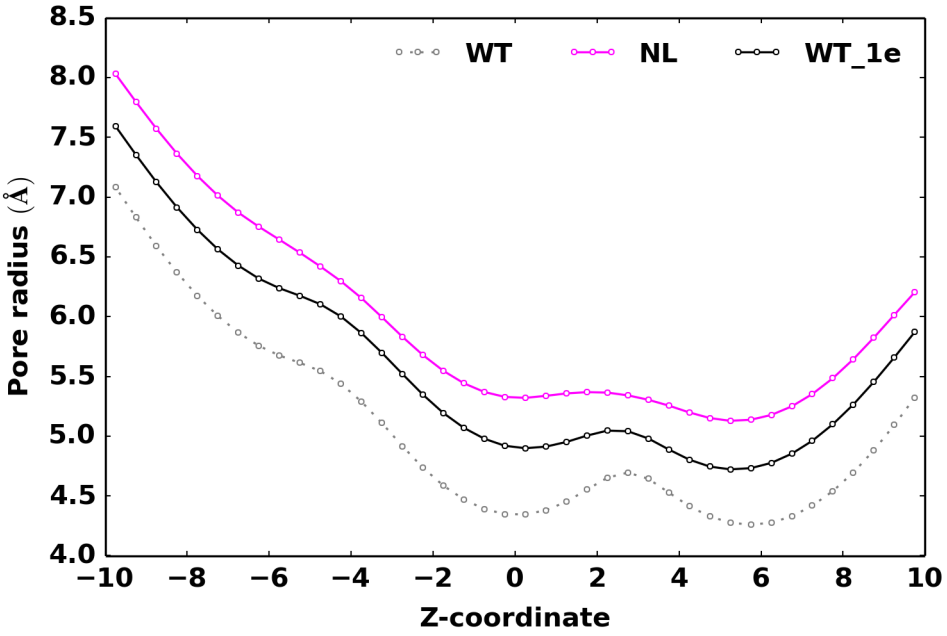


Figure 5.9: Average pore radius as a function of axial position. Average pore radius along the axial positions, obtained considering the backbone atoms.

This can be better analyzed following the time evolution of the pore radius along the axial position, as shown in Figure 5.10.

In the WT simulation, the channel is constricted around the mutated site (-1< z <1) with a pore radius of < 4.5 Å that is maintained throughout the considered time interval (Figure 5.10A), while an expansion up to 5.5 Å and 6.5 Å is observed for the WT_1e and NL models (Figure 5.10B,C). In the latter cases, a proportional expansion is observed in the regions flanking the functionalized site. The partially expanded WT_1e model was sufficient to allow ion permeation, as reported in Table 3.

Furthermore, the pore hydration upon single subunit modification was examined by monitoring the water molecules located within a defined radial distance from the channel central axis and spanning 10 Å along the z-direction on either side of position 21.

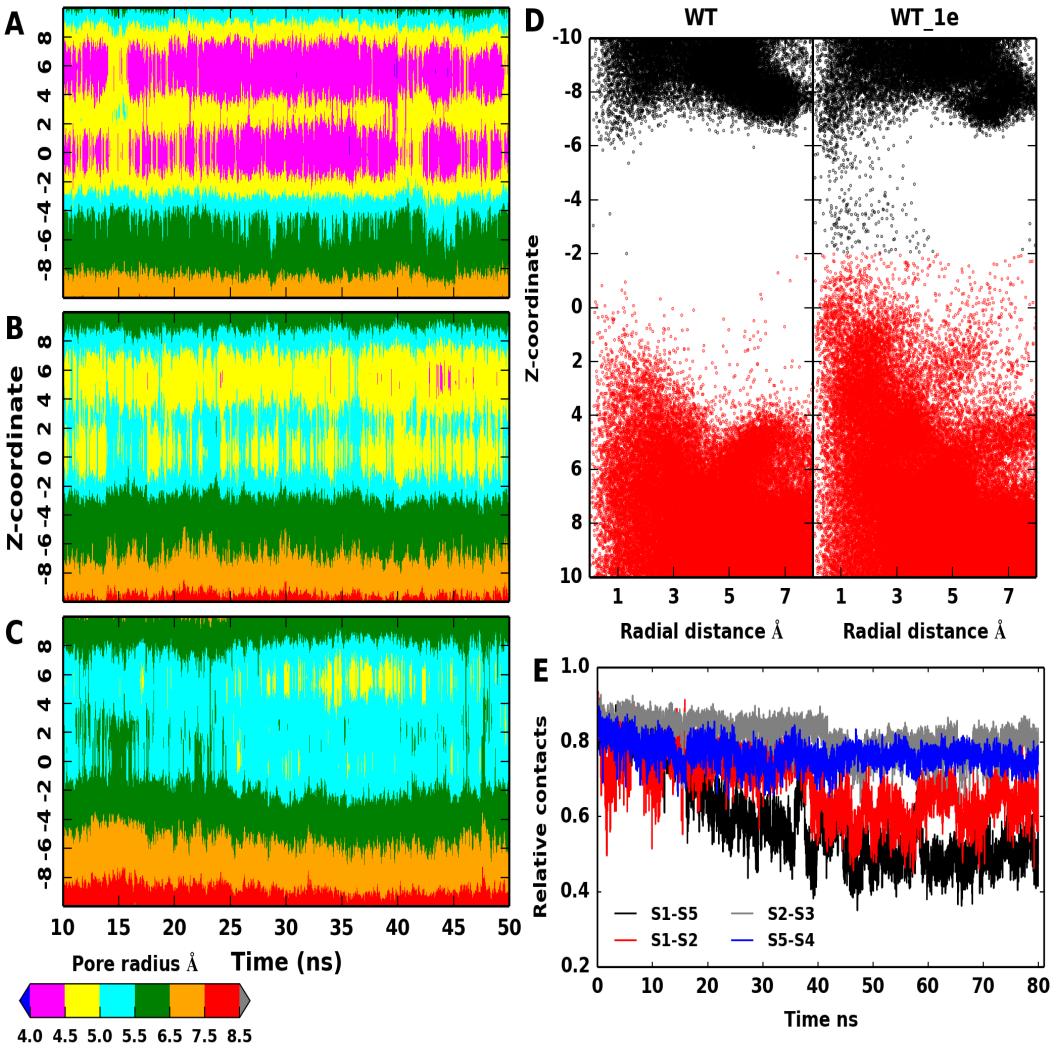


Figure 5.10: Pore radius, hydration along the channel and intersubunit contacts. Pore radius along the axial positions as a function of simulation time for (A) WT, (B) WT_1e and (C) NL models, obtained considering the backbone atoms. (D) Distribution of water molecules in the pore. For clarity, water molecules at the periplasmic vestibule of the pore are shown in black. (E) Inter-helix (TM1 vs TM1) contacts in WT_1e simulation relative to the starting configuration as a function of time.

The 2D distribution of water molecules depicted as a function of axis/radial coordinates inside the pore showed a marked difference in the hydration pattern

between WT and WT_1e models (Figure 5.10D). In the WT model, the pore is largely hydrated at the periplasmic and cytoplasmic vestibules, while the channel core region remained occluded, as expected. This occlusion extends ~8 Å and is consistent with previous estimates based on X-ray structure of Tb-MscL.[258] Introducing the charge in a single subunit (WT_1e) lessens the pore occlusion especially around the mutated site.

We also monitored the changes in inter-subunit contacts with time. In MscL, each TM1 helix does form contacts with other two TM1 helices from neighboring subunits. For example, residues of subunit 1 are in contact with residues belonging to subunit 2 and 5. We monitored the total contacts between the TM1 helix of subunit 1 (modified subunit) and TM1 helices of subunit 2 and 5 relative to the starting configuration (i.e., native contacts). The plot of variations in inter-helix contacts as a function of time is shown in Figure 5.10E. It is apparent that between TM1 helices of subunit 1 and 2, only ~50% of the native contacts are retained in the last part of the simulation, while it is further less between subunit 1 and 5 (~40%). Intriguingly, the contacts between TM1 helices of subunit 2 and 5 with their other adjacent neighbors (TM1 helix of subunit 3 and 4, respectively) are highly maintained (~80%). These observations have shown a partial detachment of the TM1 helix of subunit 1 from its neighbors, thus supporting an asymmetric rearrangement of the pore. As a result, the WT_1e channel became permeable to ions during the simulated time interval. From the number of ions passing through the pore (Table 3), we estimated a conductance of about 80 pS, smaller than the NL model but still significant.

5.3 Discussion

In the experimental context of ref. [77], where the light-responsive engineered MscL channel was originally presented, only binary scenarios were assumed: an "*off*" state with all subunits containing the neutral bulky photo-active ligands and an "*on*" state with all subunits converted to the corresponding charged moieties upon light irradiation. Here, taking inspiration from this pioneering study, we investigated in more detail through atomistic MD simulation the effect of charge activation. Starting from the "*off*" state (5L model), the photo-active ligands were sequentially replaced by charged groups and the corresponding functionalized MscL proteins thoroughly studied. Simulation results have shown a

gradual expansion of the channel proportional to the increasing number of charges in the pore lumen (Figure 5.4 and Figure 5.5), in going from the *off* (5L model) to the *on* state (NL model). In particular, the NL model displayed a conductive pore, in agreement with experiments. It is worth noting that after each chemical (electrostatic) change in the channel interior, the following structural relaxation occurred in the nanosecond timescale, thus supporting a relatively fast dynamical response of the protein channel. This result is in stark contrast to the millisecond-timescale opening observed in the native protein via the tension-induced mechanism.[259] To cope with experiments, as discussed in the following, we interpret our results as the first observation of initial MscL substates along the gating mechanism leading to the complete open channel. All systems have shown a rather stable structural character, after initial relaxation. Therefore, the observed trend in channel opening seems compliant with a gating mechanism triggered by strong charge repulsion suddenly induced in an inner region of the channel. Such an apparently simple mechanism does work so effectively only when position 21 on the TM1 helix was considered. Despite similar structural features, two alternative tested sites, namely position 17 and 25, have shown not quite the same effect, yet leading to an active MscL channel. These results are supported by several experimental evidences indicating the special role of residue G22 in Eco-MscL. Moreover, in qualitative agreement with our findings, mutagenesis studies on residues L19, V23 and G26 in Eco-MscL, which roughly correspond to L17, V21 and T25 in Tb-MscL, showed spontaneous channel activity upon charge incorporation [230,231]. In addition to such known results, in this study we attempted to provide a molecular interpretation for the effectiveness of position 21 for charge-induced activation of the MscL channel. Indeed, it has been found that this site is the most “sensitive” for the reason that it is, at the same time, the most constricted within the pore hydrophobic stretch and the less exposed to solvent. Accordingly, mutagenic studies have shown that V21 of Tb-MscL participate in the lock and reported that V21D mutant decreased the energy barrier for gating, resulting in gain-of-function phenotype.[260,261]

On the other hand, inspired by recent experiments that demonstrated channel activation upon chemical modification of a single subunit of Eco-MscL,[235,236] an additional model was considered where the charged group was introduced at only one site (i.e. V21 of subunit 1) of the Tb-MscL protein. Our simulation revealed a partially expanded pore allowing ion translocation (Figure 5.10B, Table 3), again in qualitative agreement with experiments. In contrast with the

previously described charge-repulsion trigger, the channel activation is achieved through a forced hydration into the channel core region, thus leading to pore expansion. The present gating mechanism is consistent with the proposed hydrophobic breaking effect[224] as observed with single-charge subunit substitution[235,236] or with polar residue substitution.[229] In fact, it has been observed that replacement of hydrophobic residues with hydrophilic ones triggered channel gating at reduced or no applied tension. Hence, it has been suggested that hydrophobic residues in the narrow pore region do create a functional channel block, which can be easily destabilized upon incorporation of polar (charged) residues, resulting in a conductive channel configuration through structural rearrangements.[256,257]

Our investigation has also unraveled further molecular details concerning the asymmetric motion of the only charged subunit with respect to the others. Such a structural transition, not triggered by a strong repulsive interactions, is possibly another confirm of the highly flexible character of the MscL channel. This is perhaps not surprising if we consider that the natural activation mechanism of the mechanosensitive channels does require a considerable structural change driven solely by membrane tension, whereas other protein channels activated by different stimuli (e.g., voltage and ligands) are significantly more rigid and insensitive to membrane stress. Furthermore, the MscL transition from the closed to the open state has been described as an iris-like symmetric motion of all subunits.[262-265] However, recent evidences indicated that MscL gating could also follow an asymmetric route.[266,267] In our model (WT_1e), the charge introduced in a single subunit resulted in a gradual weakening of the contacts between the latter subunit and the other pore-forming helices, while the contacts between the other native subunits have been consistently maintained (Figure 5.10E). These findings supported a progressive and asymmetric detachment of one subunit with respect to the others, an observation in agreement with experiments pointing towards a nonlinear contribution of individual subunits to the tension sensitivity.[235,236]

While the present computational investigation has provided an overall consistent picture with respect to known experiments, especially for the WT_1e and NL models representing activated MscL protein channels, some important details contrast with experimental findings. In particular, the extent of pore expansion achieved in our simulations is significantly smaller than experimental predictions. The introduction of a single charge on a pore residue in one subunit of Eco-MscL has been estimated to lead to a pore diameter of up to ~10 Å.[236]

In our work, even for the most expanded NL model, we obtained a pore diameter of about 5.5 Å around the functionalized site (including side chains). As a result, the computed conductance for both WT_1e and NL systems was found to be underestimated by a factor of about ~3 with respect to the experimental counterpart (e.g, NL model provided a conductive channel of 140 pS instead of ~500 pS[77]). We note, however, that a discrepancy between theory and experiments of the same order is not unusual,[250,268] as probably due to other important factors usually difficult to include in the molecular modeling of ion conductance, such as a more complex lipid composition, membrane pressure fluctuations, polarization effects, and so on. As an example, in a previous computational study the impact of different lipids on MscL has been addressed:[242] therein, some important structural rearrangements of the protein have been noticed, especially in the C-terminal region, though the MscL pore has resulted unaffected. In addition, a recent study showed that a specific lipid composition, including phosphatidylinositol, is crucial for allowing a physiological opening of Tb-MscL.[269]

In our case, we envisage that the expanded pore likely represents a sub-conducting state of the channel. MscL has been shown to visit sequential intermediate states during the transition from the closed to the open state. The charge-induced activation apparently populates the same initial substates achieved by the natural trigger, i.e. membrane tension.[236] Accordingly, our models presumably captured such initial substates visited upon activation whose diameter has been previously estimated to lie in the 5-10 Å range (the observed stability of the NL model channel at high temperature confirmed such a substate character). In other simulation studies focused on the natural MscL gating mechanism, a membrane tension or radial forces on protein were used to achieve an extended opening in an affordable time.[237,262,270-272] Here, we aimed at addressing the charge-triggered structural events in the absence of tension, therefore no external forces were applied to the protein or the membrane. In addition, other possible causes for the observed disagreement may be related with protein structural differences between the Eco- and Tb-MscL. Indeed, Tb-MscL has been shown to gate at about twice the tension needed to gate Eco-MscL, when expressed in *E.coli* spheroplasts (*in-vivo)* or reconstituted in azolectin liposomes (*in-vitro*), a feature that was observed when the crystals of Tb-MscL was resolubilized and functionally reconstituted.[261] Hence, despite the high conservation of the MscL pore residues among different organisms,[260]

structural features of individual MscL may have implications on its sensitivity to perturbations and interactions with the lipid environment.

5.4 Conclusions

The bacterial mechanosensitive channel of large conductance has been conceived as an ideal activated nanovalve for controlled drug delivery with several advantages that include: I) large conductance, II) non-selective pore, III) easy isolation from cells, IV) in-vitro translation, V) reconstitution in synthetic lipids and VI) charge-induced activation of channel gating. In particular, incorporation of photo-active groups into the hydrophobic channel core through site-specific mutations and chemical synthesis has enabled a light-activated gating mechanism, which offers a high spatio-temporal control without affecting the environment. In addition to these relevant features, the present computational study has unraveled a few interesting details concerning the MscL charge-induced gating: (1) a fast pore-opening dynamics (nanosecond timescale) leading to initial conducting substates, (2) a high sensitivity of the channel structural expansion to several factors (including the choice of the functionalization site in the channel interior, the number of incorporated charges, the ion distribution along the pore) and (3) an asymmetric subunit movement upon single-charge incorporation into the pore. We note that the observed sensitivity of charge perturbations towards selected channel interior sites may support the need for a preliminary molecular modeling of the detailed triggering events leading to pore gating. As well, we note that independent subunit movements upon charge activation not only support the proposed asymmetric motion in MscL channel opening but also do suggest the possibility of a subunit-specific triggering to achieve pores of desired sizes.

Chapter 6

The α-Hemolysin

Among the various different types of bio-inspired molecular devices, the bacterial PFT α-hemolysin (αHL) from *Staphylococcus aureus* represents an optimal model system, whose utility has been already tested in a large number of biotechnological applications.[78,273,274] αHL is a bacterial toxin that easily self-assembles to form an unspecific pore in the outer membranes of target cells, thus perturbing the physiological concentration gradient and leading to cell damage and death.[275] The X-ray crystallographic structure of αHL reveals a mushroom-shaped heptameric assembly composed of three structural domains: 1) the cap, constituting the exterior part of the protein, 2) the rim, which allows an effective anchoring on the surrounding lipid membrane, and 3) the stem, a β-barrel that frames the pore across the membrane (Figure 6.1A). The pore interior presents an aqueous channel that runs along the vertical axis with a diameter varying from 14 to 46 Å.[60] In a pioneering study,[69] αHL was used for the detection of small segments of single-stranded DNA, a remarkable result in view of the ongoing effort to make a cost-effective device for DNA detection and sequencing.[276] This original work has paved the way for an impressive number of applications employing αHL and similar bio-mimicking pores for molecular sensing and detection.[67,68,78] Furthermore, several features make αHL suitable for protein engineering: the overall rigid and stable structure, the internal dimension and lack of specificity, the stability under severe purification conditions, the low-noise characteristics and the effective functional assembly into synthetic lipid membranes.[277] For example, some engineering strategies proposed to exploit the αHL channel to create molecular recognition sites, e.g.

covalently attaching non-proteinogenic receptors or lodging non-covalent ring shaped adapters inside the pore to capture small analytes (drugs, organic solvents).[278]

In the present study, we report on a *de-novo* engineered photo-switch well suited for the regulation of the molecular flux across the αHL pore (Figures 6.1B and 6.1C). The molecular switch, which is designed as an extended L-shaped arm covalently linked to the αHL cap domain and equipped with a charged head group, is based on the azobenzene moiety, a widely adopted chemical group featuring a light-controllable isomerization.[279-281] In particular, this functional arm does work as an effective blocking group for channel transport by reversibly translocating in and out of the αHL pore vestibule in correspondence to the *trans* and *cis* configurations of the azobenzene fragment, respectively. Such an engineered pore has been investigated through MD simulations and free energy calculations in order to characterize the structural and dynamical properties of the photo-switch and the corresponding flux along the pore in both configurational states. Results validated the effective "on/off" inter-conversion of the switch to allow or block the passage of medium-sized molecules, such as a small DNA fragment, while the ion flux was partially affected but not fully suppressed. We believe that the present artificially gated nanopore may represent a prototype system for the rational design of a new class of nanoscale devices with controllable properties.

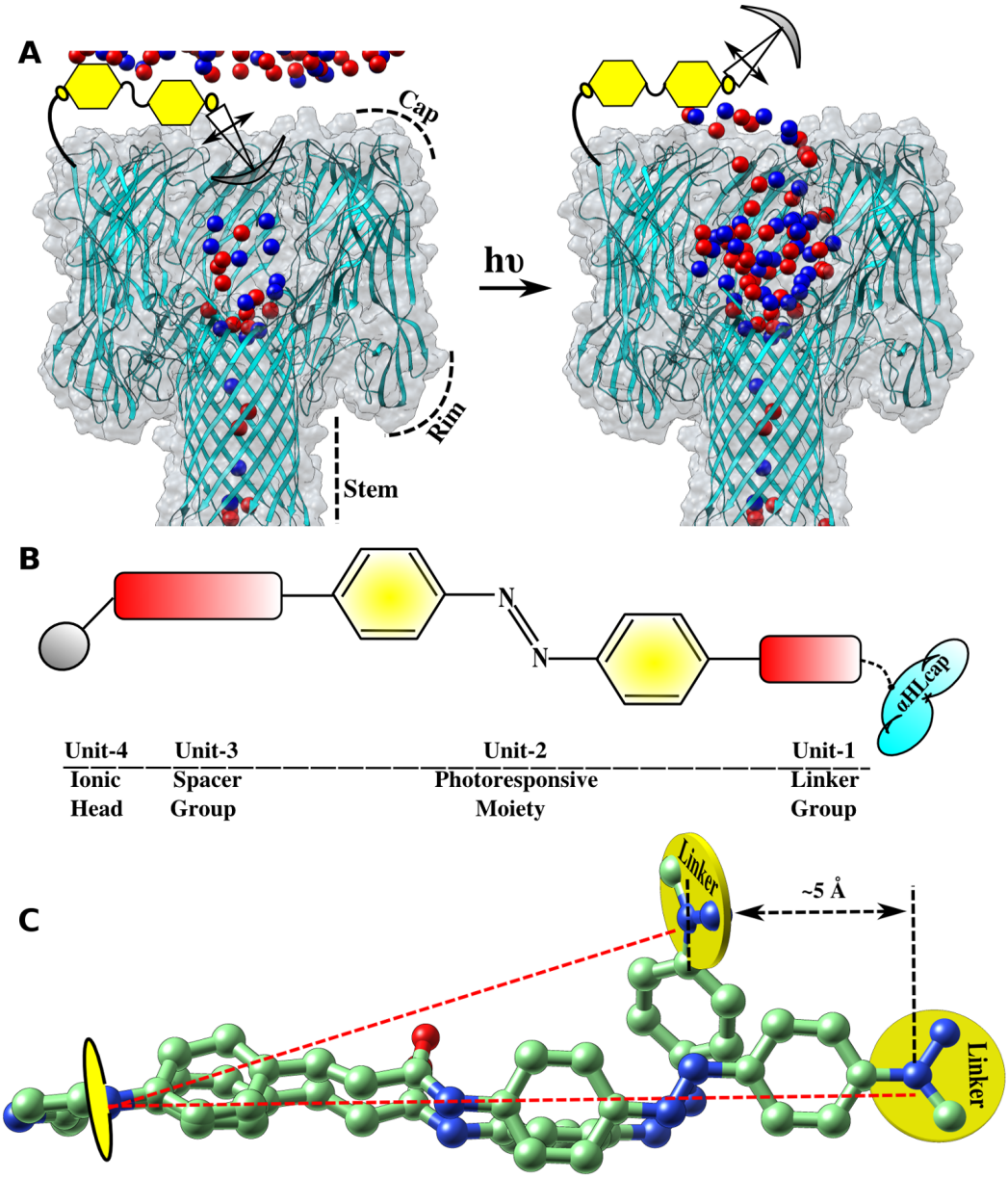


Figure 6.1: Schematic representation of the engineered αHL pore. A) α-Hemolysin (cyan ribbon) with the photo-switch attached (pictorial representation) in *trans-azo* (left) and *cis-azo* (right) configurations. Ions are represented as red (cations) and blue (anions) spheres from a given configuration. B) Scheme of the switch basic units. C) Reduction of the functional arm extension upon *trans->cis* transition.

6.1 Computational details

6.1.1 Systems preparation

The wild type αHL (PDB ID: 7AHL)[60] was embedded into a homogeneous lipid bilayer of 504 DMPC lipids (1,2-dimyristoyl-sn-glycero-3-phosphocholine) using the CHARMM-GUI server.[187] TIP3P water molecule was added up to 30 Å from the protein along with 1M KCl using the VMD software.[188] The system was equilibrated as follows: I) minimization of the whole system (5000 iterations) with an applied mild restraint (0.5 kcal/mol $Å^2$) on the protein C_α coordinates, II) gradual heating from 10 K to 300 K (about 0.5 ns) and III) equilibration at constant pressure (1 atm) and temperature (300 K) for about 12 ns. Afterwards, a production run was performed and a molecular model of the switch was generated with the azobenzene in the *trans* configuration (hereafter, *trans*-azo), following the scheme depicted in Figure 6.2 (see also Sec. 3.1 for more details), and covalently attached to residue 237, which is located at the cap domain exterior (Figure 6.1A). Neutralizing counter ions were added and the total concentration was readjusted to 1M. After a short equilibration (500 ps in a NPT ensemble), a production run was performed. Hence, from a well-equilibrated configuration of the *trans*-azo model, an independent production simulation was initiated by adopting force field parameters corresponding to the *cis* configuration of the azobenzene fragment (hereafter, *cis*-azo).

6.1.2 Simulation and analysis details

Molecular dynamics. MD Simulations of WT, *trans*-azo and *cis*-azo systems were performed in a NVT ensemble, applying a homogeneous external electric field along the z-axis (perpendicular to the lipid bilayer) corresponding to a defined voltage (1 Volt), for about 50ns each. All simulations were performed under periodic boundary conditions with the NAMD code (v 2.9) using the CHARMM27 force fields for the protein, lipids[190,245,246] and the general small molecule force field.[249] The azobenzene force field parameters were obtained from previous studies, describing a flexible molecular model for both isomers.[282,283] Atomic charges for the azobenzene fragment were obtained

following the restrained electrostatic potential (RESP) method to fit the quantum-mechanically derived molecular electrostatic potential from the B3LYP/6-31G* calculations, using Gaussian09[247] and R.E.D[248] software (see Figure 6.3 and Table 5.1). The parameters for the remaining portion of the functional arm were adopted from the CHARMM CGEN general small molecule force field.[249]

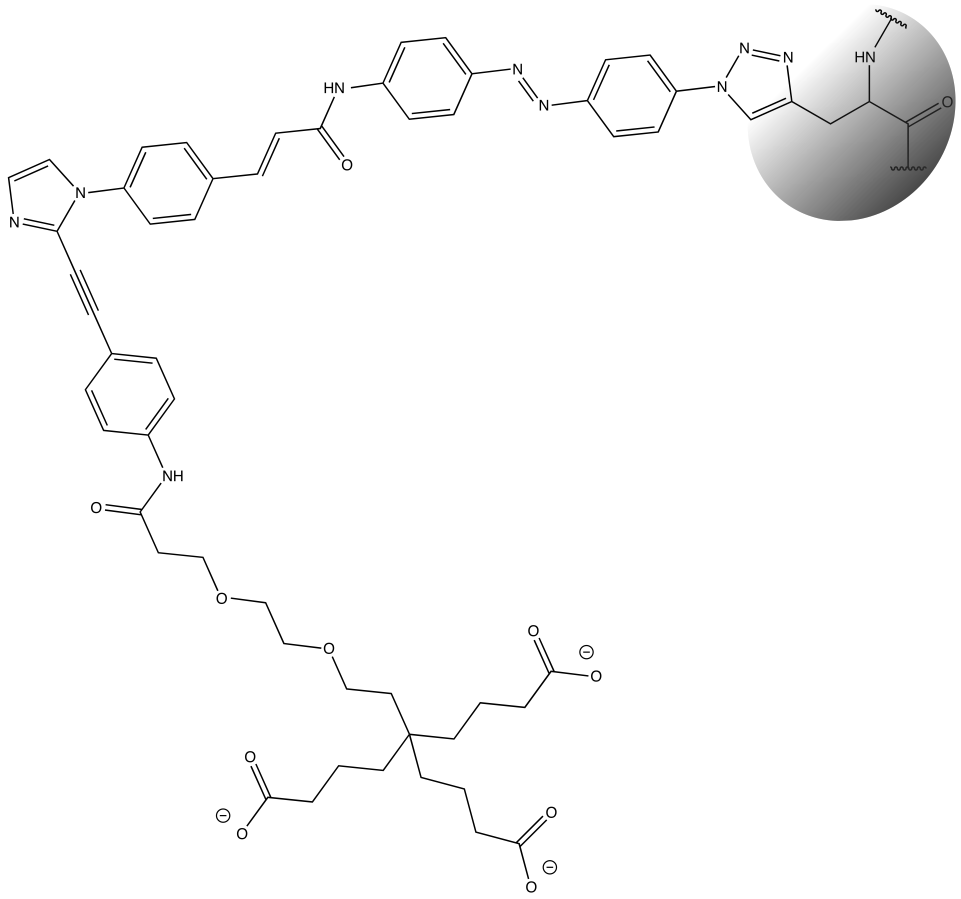


Figure 6.2: Scheme of the engineered photo-switch chemical structure. The gray region embeds the atoms belonging to the non-natural residue 237.

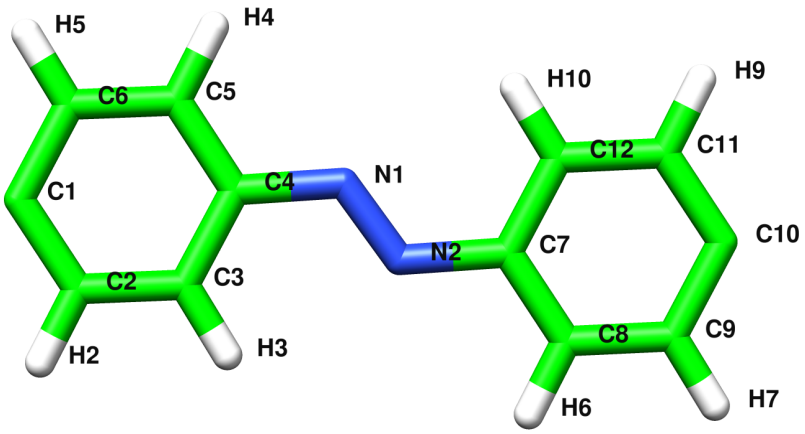


Figure 6.3: Structure of the azobenzene molecular model.

Atom	Trans-azo	Cis-azo	Atom	Trans-azo	Cis-azo
C1	-0.0131	0.0458	N2	-0.123	-0.2922
C2	-0.1699	-0.1926	C7	0.1391	0.4458
H2	0.1511	0.1572	C8	-0.0638	-0.2533
C3	-0.0638	-0.2533	H6	0.0811	0.189
H3	0.0811	0.189	C9	-0.1699	-0.1926
C4	0.1391	0.4458	H7	0.1511	0.1572
C5	-0.0638	-0.2533	C10	-0.0131	0.0458
H4	0.0811	0.189	C11	-0.1699	-0.1926
C6	-0.1699	-0.1926	H9	0.1511	0.1572
H5	0.1511	0.1572	C12	-0.0638	-0.2533
N1	-0.123	-0.2922	H10	0.0811	0.189

Table 1: Atomic charges of the azobenzene moiety in the trans-azo and cis-azo system.

Bonds involving hydrogen atoms were kept rigid using SHAKE,[284] with a 2 fs timestep for the numerical integration of the equations of motion. Non-bonded interactions were calculated with a smooth cut-off (from 10 to 12 Å). Long-range electrostatic interactions were evaluated with the PME method.[107]

Constant temperature was enforced (300 K) using a Langevin thermostat.

Ion flux through the pore was analyzed by counting the number of permeated ions with an in-house code written using MDAnalysis.[196] Pore conductance was estimated from the ratio of the average current and applied voltage.[250,251] Average current was estimated from a linear fit of the cumulative current obtained by integrating the instantaneous current computed as,

$$I(t) = \frac{1}{\Delta t \cdot L_z} \sum_{i=1}^{N} q_i \left[Z_i(t + \Delta t) - Z_i(t) \right] \qquad (1)$$

where q_i and Z_i are the charge and z-coordinate of the i-th ion, respectively, Δt is a parameter set to 5 ps and L_z is the protein length in z-dimension. Figures and plots were generated with UCSF-Chimera and Matplotlib tools.[197,198]

Steered MD and free energy calculations. A single stranded polynucleotide chain (poly-dA), comprising six nucleotides, was pulled through the engineered αHL pore (both *trans*-azo and *cis*-azo systems) using constant-velocity steered molecular dynamics (SMD). According to this methodology, the molecule of interest is attached to a dummy atom via a virtual spring and the force experienced by the pulled molecule is evaluated as the dummy atom moves along a predefined reaction coordinate with constant velocity.[285] The poly(dA) was generated using the Chimera software and oriented in such a way to align its principal axis along the pore axis. The poly(dA) center of mass was aligned with the geometric centroid of the C_α atoms belonging to residue K21 on each protein monomer and shifted by 10 Å along the z-direction to place the molecule at the cap domain exterior. Water molecules overlapping with the inserted polymer were removed and counter ions were added to keep the whole system electrically neutral. The system was equilibrated for 5 ns in a NPT ensemble, restraining the poly(dA) backbone atoms along with the protein C_α atoms. Then, poly(dA) was pulled towards the pore interior using a rate of 2 Å/ns and a spring force constant of 5 kcal/mol $Å^2$.

The PMF driving the insertion of the arm inside the pore vestibule was computed using the ABF methodology implemented in NAMD.[126,128] To this end, a predefined coordinate corresponding to the distance has been adopted. The average force acting on the head group along such a coordinate was estimated

during the simulation and an adaptive biasing potential, equivalent and opposite to the applied average force, was derived. Ten distinct simulations were performed according to spatial windows of 2 Å length, spanning an extended region within the cap domain, each one for about 28 ns. Within each window the average force acting on the selected ion was accumulated in 0.1 Å sized bins, and the biasing force was applied after 800 samples. Starting coordinates were extracted from an equilibrated configuration of the production run described above.

6.2 Results

6.2.1 Description of the artificial light-driven switch

The photo-switchable molecular arm, which confers an artificial gating to the αHL protein channel, was conceived on the basis of a few simple designing principles, as described in the following. It is worth noting, however, that many other structures could have been possibly proposed. Hence, the specific details of the arm molecular architecture should be regarded as a matter of convenience, serving the main purpose of this work.

The chemical structure of the functional arm, as tagged to the protein, is reported in Figure 6.2. First, we selected the azobenzene moiety as the photo-sensitive trigger, since it is one of the best-known molecular switches, owing to its simple chemical structure, fast intramolecular conversion and stability over multiple photo-isomerization reactions. Azobenzene exhibits two isomeric forms: the extended planar *trans* (hereafter, *trans*-azo) and the twisted *cis* (hereafter, *cis*-azo) configurations. The *cis*-*trans* photo-isomerization is known to result in an increase of the distance between the para-carbon atoms of about 3.5 Å.[273,286] Second, we devised a rigid and approximately linear molecular framework in order to maximize the distance between the edges of the two isomers: the resulting reduction in arm length while going from the *trans* to the *cis* configuration was about 5 Å (Figure 6.1C). In addition, the rigid structure was conceived to restrict the functional arm movements and, in turn, to favor a fast structural transition from/to the pore. The engineered arm was also equipped with a 1-phenyl-2-(phenylethynyl)-1*H*-imidazole moiety, so as to introduce a “kink” into the linear framework, approximately at the center of the pore entrance,

orienting the arm head group versus the channel interior. Third, we considered a residue (i.e., position 237 on one subunit) on the αHL cap domain for covalently attaching the functional arm, with the aim to obtain in the "open" state a totally unobstructed pore, resembling as much as possible the wild-type channel in terms of transport properties. In fact, even if the azobenzene moiety could be attached inside the pore vestibule, as shown in ref. [75], this would lead to a significantly narrower pore, possibly preventing large-compound translocation (e.g., poly-peptides and poly-nucleotides). Concerning the protein-arm linking, we considered a recent approach exploiting the site-specific substitution of a residue with a non-natural amino acid carrying a side chain (here, an alkyne group) suitable for bio-orthogonal chemistry:[287] upon cycloaddition reaction with an azide group, the substituted residue forms a 5-membered ring, namely 1,2,3-triazole (Figure 6.2). In particular, a single monomer of the α-Hemolysin channel could be modified and assembled to form a hetero-heptamer by following chemical and co-expression techniques (followed by a purification step) as shown recently for α-Hemolysin[75] and MscL[235]. Finally, the arm head group has been designed to be both flexible and negatively charged by three branched butyric acid groups, so to drive the arm into the pore under the action of an external electric field, i.e. under usual experimental conditions for nanopore applications. In particular, the flexible aliphatic ether linkers have been introduced to let the arm head occupy an extended portion of the pore vestibule, thus effectively hampering molecular transport through the channel.

6.2.2 Blocking and unblocking of the engineered αHL pore

To simulate the experimental scenario of both isomeric configurations as issuing from the photo-isomerization reaction, independent MD simulations were performed on the engineered αHL protein according to the *trans-azo* and *cis-azo* models (see Sec. 6.1 for further details). Note that the *cis-azo* system was simulated from a previously equilibrated configuration of the *trans-azo* system, hence in the following results it is apparent the relaxation of the former from a non-equilibrium state. First, we have analyzed the molecular arrangement of the arm head group with respect to the pore vestibule. As a relevant parameter for the present analysis, we have adopted the minimum distance between the centroid of each carboxyl group and the centroid of the C_α atoms belonging to residue K21,

which identifies a convenient reference point at the center of the αHL cap domain (depicted as a yellow sphere in Figures 6.4C and 6.4D).

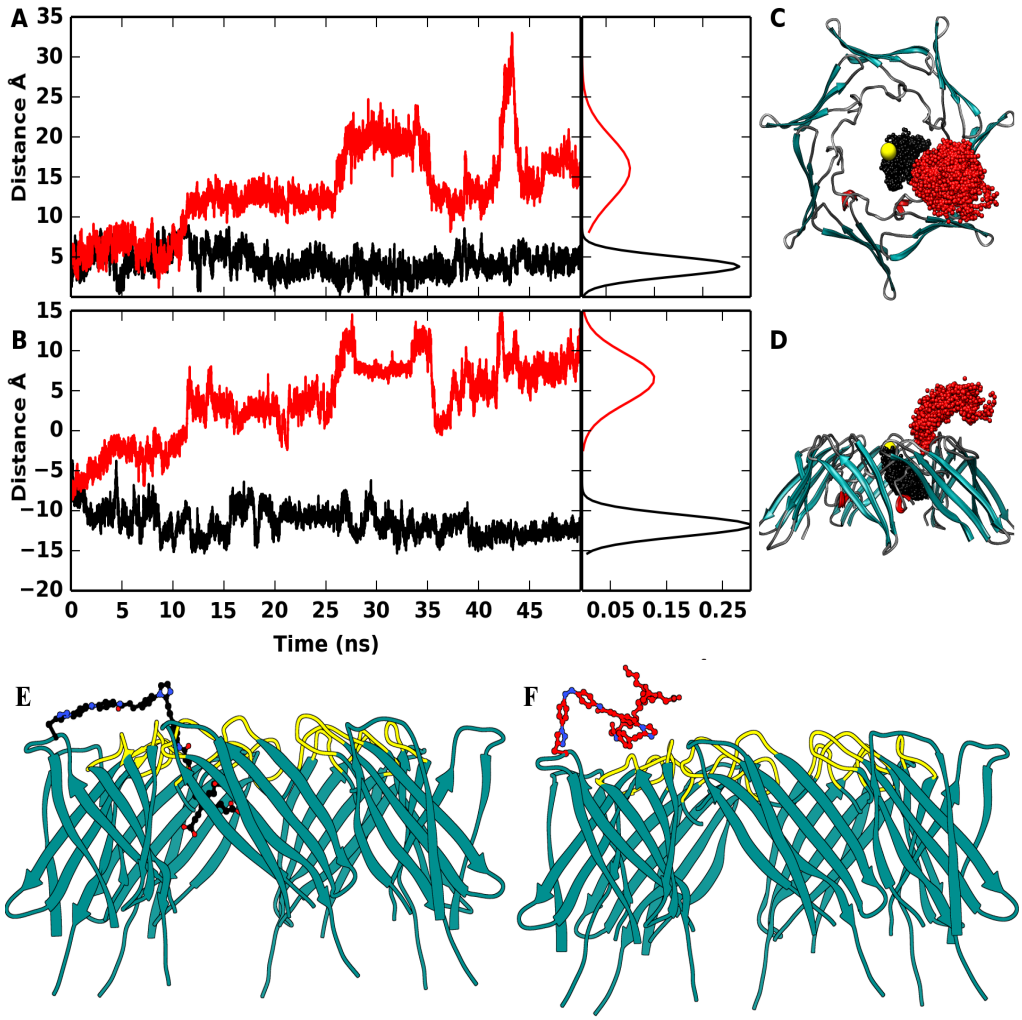


Figure 6.4: Structural fluctuations of the photo-switch in both isomeric forms. A) Minimum lateral distance of the head charged groups from the channel centroid (in the xy-plane) along with the corresponding distribution over the last 30 ns. B) Minimum axial distance to the channel centroid (z-axis) and corresponding distribution. Data for the *trans-azo* and *cis-azo* systems are shown in black and red, respectively. Spatial occupancy of the functional groups from top C) and side D) views (the yellow sphere indicates the C_α geometric centroid of K21). Representative structures of the functional arm in the E) *trans-azo* and F) *cis-azo* configurations (in yellow, residues shaping the pore entrance).

In Figure 6.4, the time evolution and the distribution of the projections of such a distance along both the lateral (xy-)plane and the longitudinal (z-)axis of the pore are reported.

A clear difference emerged between the *trans-azo* (black line) and the *cis-azo* (red line) simulations. Concerning the projection along the lateral plane (Figure 6.4A), an average distance of ~4.5 Å, with a rather narrow distribution, was obtained with the *trans-azo* model, which was maintained throughout the simulation. On the contrary, in the case of the *cis-azo* system, the distance drifted to about 12-15 Å after 15 ns, showing large fluctuations afterwards (up to 30 Å). An even starker difference was obtained evaluating the displacement along the longitudinal axis. The *trans-azo* configuration featured an average distance of about 13 Å below the reference point (Figure 6.4B), according to a stable arrangement of the arm within the pore interior (i.e. closed form). At variance, the *cis-azo* model led to a distance distribution well displaced above the reference point, reaching an axial distance up to 15 Å, in agreement with a gradual motion of the head group away from the pore vestibule (i.e. open form). Note that the two average protein structures were instead rather similar due to the application of the same atomic restraints, as shown in Figure 6.5. Besides, the observed marked difference between the two configurational states was also visualized through a 3D representation of the head group spatial occupancy with respect to the cap domain, as shown in Figures 6.4C and 6.4D: from the spatial distribution it is apparent the placement of the artificial arm in the exterior region of the cap domain according to the *cis-azo* model. Note that residue 237 backbone dihedrals confer enough flexibility to allow the arm rotation with respect to the cap domain in the *cis-azo* system (see Figure 6.6).

Furthermore, we have examined in some detail the distribution of the carboxyl groups while the arm was anchored inside the pore (*trans-azo* model). In this case, the centroid of each carboxyl moiety was monitored as a function of time and its distribution reported in Figure 6.7: results showed that the three charged groups freely explored a wide area within the pore vestibule, suggesting the possible alteration of the molecular flux (e.g., ions, molecules) across the pore in this configuration.

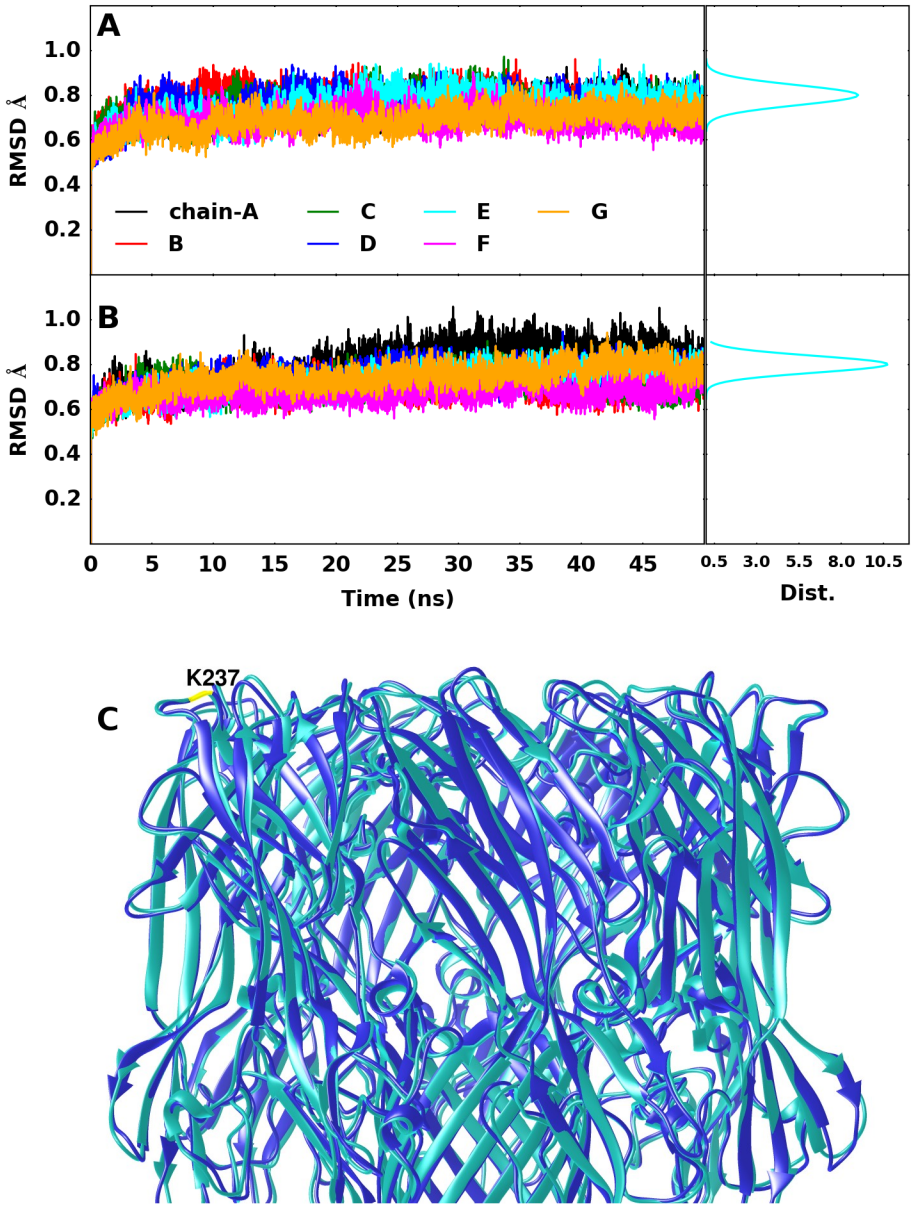


Figure 6.5: Comparison of Hemolysin protein structures. Backbone RMSD of each individual subunit of Hemolysin as a function of time for (A) *trans-azo* and (B*) cis-azo* systems. The distribution of the RMSD values for the modified subunit (chain E) is depicted in the right panel. (C) Superimposition of the average protein structures over the last 5 ns as obtained from both simulations.

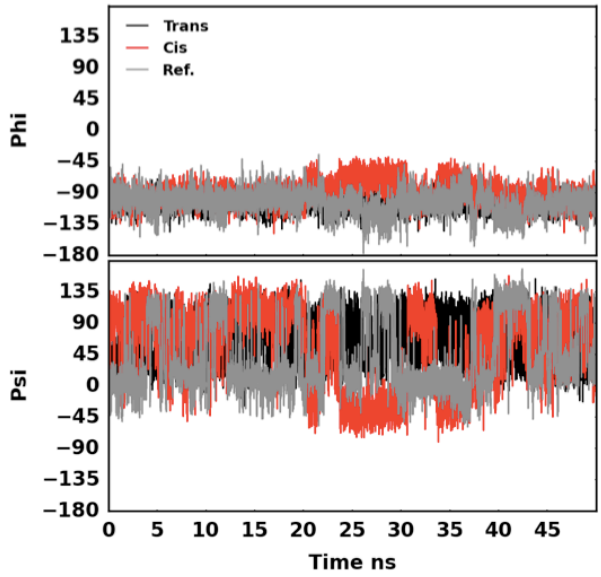


Figure 6.6: Structural flexibility of residue 237. Time evolution of residue 237 Phi (top) and Psi (bottom) angles as a function of simulation time for (black) trans-azo, (red) cis-azo and (grey) native protein.

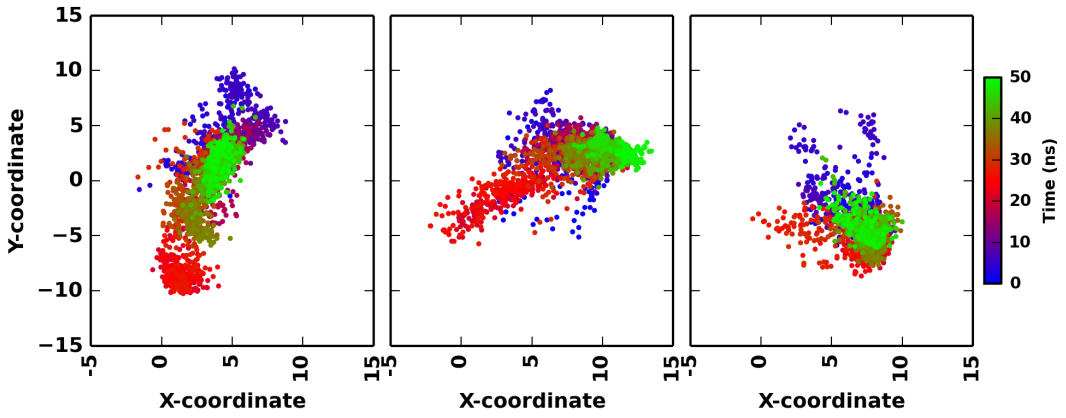


Figure 6.7: Lateral distributions of the carboxyl groups within the pore vestibule. Distribution of the carboxyl groups along the plane perpendicular to the channel axis as a function of time, according to the *trans-azo* system. Data for each carboxyl group are presented in a separate panel.

To gain further information about the energetic stability of the *trans-azo* system in the apparent favorable configuration within the pore interior, we have performed a free energy analysis of the latter system along an exit pathway from the equilibrium closed form. To this end, we have selected, for convenience, a simple "reaction coordinate" given by the relative distance between the head group and the pore center (i.e. the reference point defined above) and we have evaluated the potential of mean force along this coordinate. Not surprisingly, the PMF profile has shown a distinct energy minimum in correspondence to the equilibrium structure of the *trans-azo* arm (about 5 kcal/mol, around z = -14 Å) and an energetically expensive pathway towards the pore exit, as depicted in Figure 6.8. Conversely, the computed PMF showed a downhill from the protein extracellular side versus the channel interior, thus supporting the easy insertion of the *trans-azo* arm into the pore.

Moreover, to investigate the reverse *cis->trans* interconversion and the possible translocation of the functional arm inside the pore, we have selected a few equilibrated molecular configurations of the *cis-azo* model and performed additional simulations by imposing the force field parameters characterizing the *trans-azo* system. Note that the main purpose here was to study the structural transition of the functional arm with respect to the pore, not the isomerization reaction itself. In one case, we have observed a successful insertion of the ionic head into the pore vestibule within the spanned time interval (the insertion was completed in about 30 ns), as revealed by the distance of the latter with respect to the same reference center (bottom panel, Figure 6.8). Once fully translocated into the pore, the functional arm remained trapped in this "closed" state. In other cases, a clear translocation of the engineered arm into the αHL channel was not observed within the accessible simulation time (a few tens of nanoseconds), being the overall kinetics, as expected, largely dependent on the initial configuration. However, despite the unpredictable kinetics, our results supported the unidirectional translocation of the switch charged head into the pore, upon *trans* activation.

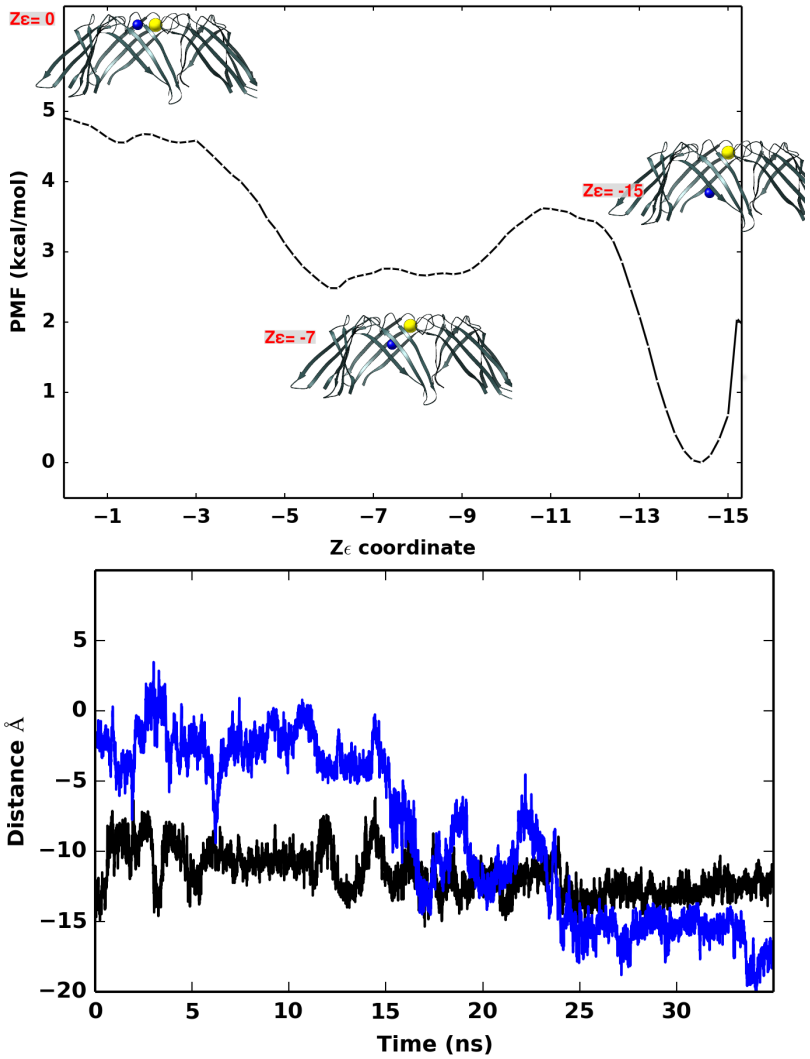


Figure 6.8: PMF profile of the head group translocation into the pore. (Top panel) PMF profile of the head group translocation as a function of the z-coordinate. The origin was set at the C_α geometric centroid of residue K21 at the cap domain vestibule. The location of the functional head (blue sphere) with respect to the reference coordinate (yellow sphere) is schematically depicted along with a relevant protein region (cyan). (Bottom panel) Minimum axial distance between the charged head groups and the channel centroid for *trans-azo* model: in blue, initial configuration taken from *cis-azo* system; in black, equilibrated *trans-azo* system.

6.2.3 Ion permeation and pore conductance

MD simulations, in combination with applied electric fields, are frequently used to estimate ionic flux and conductance in biological channels.[249,251] However, due to the molecular complexity of the systems under scrutiny, we focused on a comparative analysis of ion permeation and conductance with no attempt to reproduce experimental observables. In particular, we estimated the ionic current across the pore by following two different methodologies, one based on simple ion counting and the other based on the overall motions of the ions along the pore direction. First, the ionic flux across the engineered pore was estimated by evaluating the number of ions entering into the pore (both the number of ions just entered from one side of the pore, either the cap or the stem domain, and the permeated ions, those translocated across the pore). Results of the engineered pore in both configurations are reported in Table 5.2 along with those issuing from the WT protein.

System	Tot. Entry[a]	Perm.[b]	Rel. Perm.[c]
WT	338	250	1.00
Trans-azo	288	194	0.78
Cis-azo	329	257	1.03

Table 2: Summary of ion translocations through the WT and engineered pore. [a)]Total number of ion entries, [b)]number of fully permeated ions and corresponding [c)]relative permeation with respect to WT system, issuing from the last 30ns of simulation.

The obtained statistics indicated a somewhat lower percentage of permeated ions across the *trans-azo* model with respect to the WT and the *cis-azo* model, about 22% considering both K^+ and Cl^- ions. In the *cis-azo* case, no appreciable difference (~3%) with respect to WT was obtained. Then, the cumulative current through the WT and the engineered pore was derived from the instantaneous current computed over the last 40 ns of simulation. As depicted in Figure 6.9, the

computed current displayed an apparent divergence of the *trans-azo* model with respect to the other systems after about 25 ns: a further indication of a larger flux across the WT and the *cis-azo* pore. As a result, the estimated pore conductance for the WT, *cis-azo* and *trans-azo* systems was 1.88, 1.80 and 1.60 nS, respectively. Hence, the conductance of the *trans-azo* model was predicted to be lower by ~15% with respect to the other systems, in qualitative agreement with the ionic flux analysis. Overall, these results confirmed the influence of the anchored ionic head on the flux across the pore.

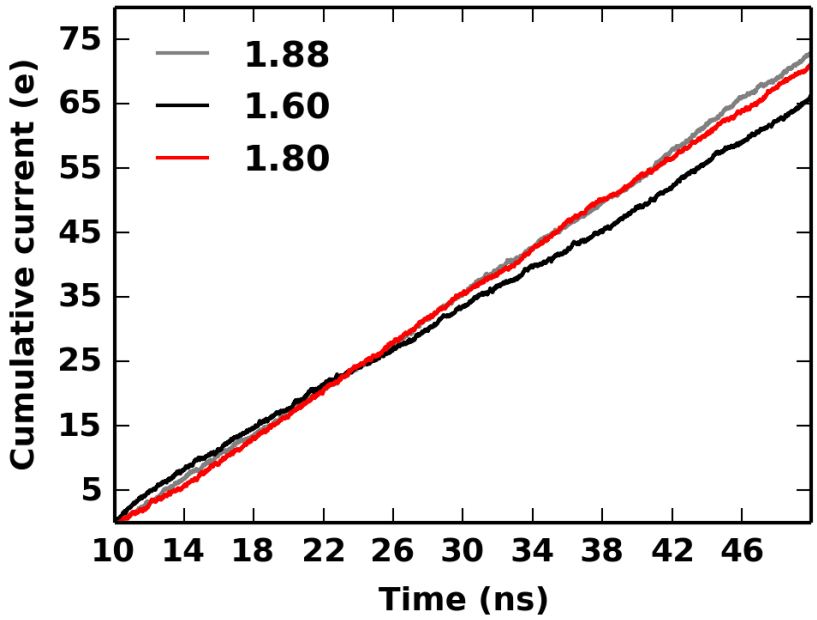


Figure 6.9: Cumulative current as a function of simulation time. Cumulative current (in unitary charge) issuing from the WT (gray), *trans-azo* (black) and *cis-azo* (red) systems over the last 40 ns of simulation. Legends indicate the estimated ion conductance in nS.

6.2.4 Translocation of poly(dA) through the engineered pore

The translocation of a single-stranded DNA fragment, namely $(dA)_6$, through the engineered αHL was carried out exploiting a computational approach similar to previous studies.[285] In particular, we applied the constant-velocity SMD technique, as described in Sec. 2.2. The poly(dA) was pulled down, from the cap

domain vestibule to the pore interior, along the pore longitudinal axis (Figure 6.10A). Accordingly, the force exerted by the pulling spring connecting poly(dA), on one side, and the moving ghost particle, on the other, was monitored as a function of time, as reported in Figure 6.10B.

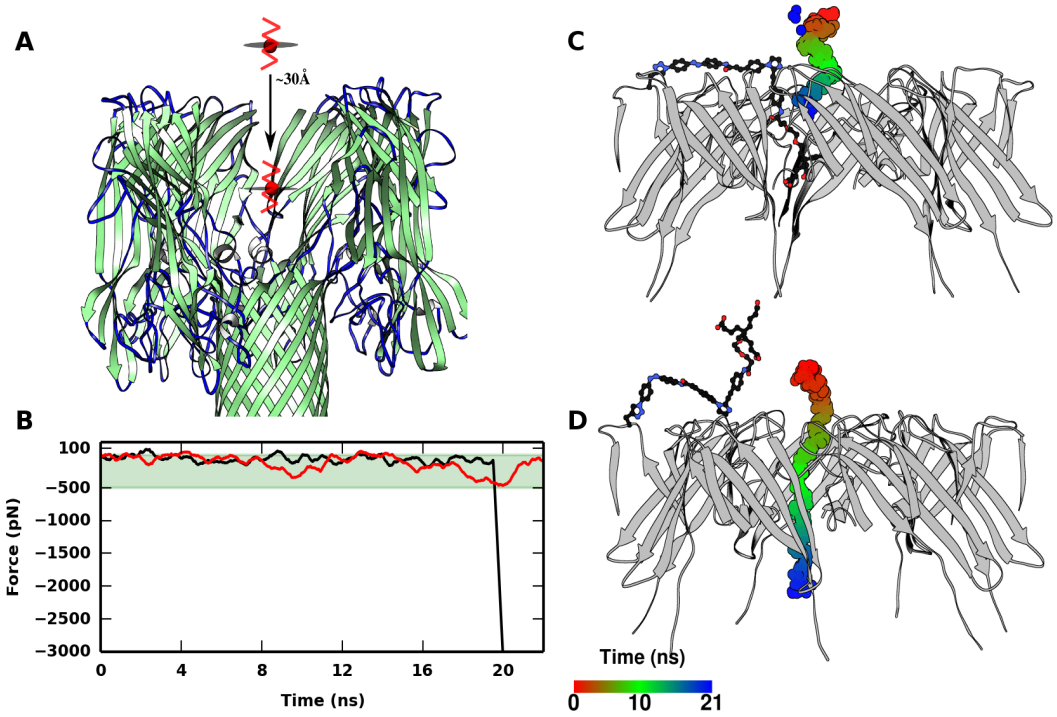


Figure 6.10: Poly(dA) translocation through steered MD simulations. A) Poly(dA) pulled down from the exterior cap domain into the pore, covering a distance of ~30 Å along the channel axis. B) Pulling force as a function of time: in black, *cis-azo* system; in red, *trans-azo* system. Representation of the poly(dA) translocation through the C) *trans-azo* and D) *cis-azo* models, depicted as a function of time. The poly(dA) center of mass is represented as a sphere whose color is proportional to the elapsed time.

The profile of the computed force showed smooth fluctuations in both engineered systems up to ~20 ns of simulation. Beyond such a time interval, the computed force displayed a sharp increase in the *trans-azo* model, while it remained approximately similar in the case of the *cis-azo* model. The different behavior was also evidenced by visual inspection of the two systems: the time evolution of the poly(dA) center of mass was traced inside the pore, as shown in

Figures 6.8C-D. In the *trans-azo* model, the well-anchored head firmly obstructed the passage of $(dA)_6$, which could not translocate any further along the pore after the initial insertion (Figure 6.10C). Indeed, the negatively charged carboxyl groups exerted an effective repulsive interaction, along with other possible steric effects, towards the translocation of the molecule. On the other hand, in the *cis-azo* model, the absence of any physical obstacle easily allowed the transport of the DNA polymer chain through the αHL channel (Figure 6.10D), similarly to what was observed experimentally with the native αHL channel.[69]

6.3 Discussion

Our *in-silico* investigation has shown how the proposed engineered photo-switch can be effectively integrated with the well-known αHL membrane channel to serve as a controllable molecular "anchor" in various nanopore applications where αHL can be fruitfully exploited. The present molecular design followed a few basic prescriptions aiming at enhancing the effect of the azobenzene structural change upon illumination. Indeed, we demonstrated the reversible dynamics of the functional arm in response to the photo-isomerization reaction of this moiety, within the approximations of the model. Our results, based on atomistic MD simulations and free energy calculations, consistently showed an effective and reversible structural transition of the engineered αHL pore between a "blocked" state, corresponding to the equilibrium configuration of the *trans-azo* model, and an "unblocked" state, corresponding to the *cis-azo* configuration. We have investigated both the *trans* -> *cis* and *cis* -> *trans* inter-conversions, in which the functional arm underwent a large amplitude motion driven by the concomitant effects of the azobenzene rearrangement and the applied voltage. The more extended configuration characterizing the *trans-azo* model allowed a stable placement of the ending portion of the functional arm within the pore vestibule, while the applied electric field further facilitated the penetration of the head group down the channel. Free energy analysis provided further information about the energetically favorable insertion pathway of the arm, according to the *trans* configuration: the PMF featured a smooth downhill, thus supporting a funnel-like shape of the free energy surface describing the head group translocation inside the channel. Conversely, in the twisted *cis-azo* geometry the arm length was substantially reduced by about ~5 Å (See Figure 6.1C) and kinked at the

azobenzene unit by about 80°. As a consequence, the functional arm could not fit comfortably into the αHL channel, as seen in the *trans* form, and its charged head group was observed to gradually exit from the pore vestibule. Therefore, the switch successfully dropped and retreated a molecular "anchor" within the large pore of the αHL protein. We have not pursued an accurate description of the inter-conversion kinetics, since it would have required the challenging treatment of both reactive and long-time scale events. Nevertheless, we may reasonably predict a faster transition from the *closed* to the *open* state (in the order of tens of nanoseconds) with respect to the opposite route (i.e., *cis* -> *trans*), owing to the large ensemble of structures characterizing the *cis-azo* model.

The stark difference of the two configurational states previously described had important consequences on the transport properties of αHL. The charged moieties of the functional arm in the anchored geometry (i.e., *trans-azo* model) altered, though to a minor extent, the ionic current and the corresponding conductance with respect to the unanchored configuration (i.e., *cis-azo* model). The latter state, on the other hand, resembled closely the conductive properties of the WT system, a significant result implicitly pursued since the original design of the functional arm. Hence, the computed ionic conductance of the two states differed by a small fraction (about 15%), nevertheless potentially useful for probing the status of the pore in single-channel recording applications. The most important acquired functionality of the engineered αHL channel, however, was the capability to totally block the translocation of molecular species, as shown by the illustrative case of a small polynucleotide chain. Remarkably, the equilibrium *trans-azo* configuration featured an impenetrable barrier along the channel pathway for the transport of poly(dA) (Figure 6.10), whereas no obstruction was present upon *trans->cis* transition, in which case the αHL pore recovered its natural transport capabilities. The barrier was ascribed to the combined effect of steric hindrance, being the pore vestibule partially occupied by the engineered arm, and electrostatic interactions exerted by the charged head.

The relevance and applicability of a photo-switchable biological pore can be better assessed by taking into consideration similar approaches developed so far. First, we note that the αHL pore was used in a number of applications in which pore blockage was pursued through appropriate obstructing molecules.[273,274] On the other hand, the photo-switchable properties of the azobenzene moiety were exploited in an even larger set of applications.[279-281] In particular, it is interesting to comment on the photo-switch mechanism proposed in this study as

compared to two recent applications exploiting the azobenzene light-trigger in connection with the αHL protein. Previously, Bayley et al. conjugated a sulfonated azobenzene at position 117, located in the pore lumen along the stem domain, and showed that the *trans* isomer could affect the ionic conductance of the pore appreciably more than the *cis* counterpart.[75] More recently, a photochromic host-guest supramolecular system was implemented to reversibly regulate the open/close states of αHL.[288] The sulfonato-calix[4]arene (SC_4) molecule was attached to lysine residues at the stem entrance via host-guest interaction, thus showing the effective closure of the pore. Later addition of trans-azobenzene viologen guest molecule displaced the SC_4 from the stem, favorably competing with lysines towards SC_4 binding, thereby leading to the pore opening. On the contrary, the corresponding cis-azobenzene counterpart inhibited the αHL pore opening. Such a notable application was shown to successfully lead to a reversible "on/off" gating mechanism, though it required the concerted action of a multi-component system and its transition rate was diffusion limited. Altogether, previous applications have inspired and supported the original protein design outlined and tested in the present study. However, at variance with the two approaches described above, the photo-switch proposed here works at the level of the cap domain, which frames a large solvent-filled channel, and requires no addition of further chemicals. This choice offers a fast-responsive control of the channel transport properties, not perturbing the pore functional assembly in the lipid membrane environment. Besides, in view of analytical applications, the photo-anchor can be further exploited for lodging appropriate molecular receptors and adapters inside the pore, so to introduce, in principle, specificity and selectivity towards molecular transport.

6.4 Conclusions

Our study highlighted the possible modulation of ion conduction and molecular transport across a purposely engineered αHL protein by a *de-novo* designed light-controllable molecular switch. From a thorough computational investigation, we have shown how such a bio-inspired artificial system, exploiting the photo-isomerization of the azobenzene moiety, has successfully introduced a smart "nanovalve" into an unspecific non-gated biological pore. Results evidenced two limiting configurational states corresponding to the "on" and "off"

status of the photo-switch. While the open state, namely the *cis-azo* configuration, displayed transport properties in very good agreement with the natural αHL protein channel, the closed state (i.e., the *trans-azo* model), characterized by the stable insertion of an electrically charged synthetic group into the pore vestibule, showed remarkably different properties. In our study, we tested the translocation of simple ions as well as a single-stranded DNA fragment, as illustrative examples of relevant systems for nanopore applications. Indeed, we have observed a significant reduction of the ion conductance and the effective inhibition of the DNA translocation upon transition to the artificially closed configuration.

We believe that such a prototype system may support the design and synthesis of molecular devices with suitable light-driven functions for nanopore applications at single-molecule level. For example, the present approach can be tailored to introduce photoresponsive capabilities in other unspecific and non-gated biological pores such as ompG, MspA etc. Alternatively, the present molecular switch can be modified for addressing selective transport and delivery by tuning the chemical nature, the size and/or the charge of the functional group. Further work will be focused on the systematic assessment of the transport properties through pore engineering.

Chapter 7

The IscS-IscU complex

The ISC (Iron Sulphur Cluster) system is one of the three main [Fe-S] protein biogenesis systems known in bacteria[289] and is responsible for the maturation of "housekeeping" [Fe-S] proteins under normal physiological conditions. Recently it has been discovered that an increasing number of human pathologies are associated to defects in ISC system proteins,.[95,290,291] Amongst these are: Friedreich's ataxia (MIM 229300);[292] the myopathy with lactic acidosis and exercise intolerance (MIM 255125);[293-295] childhood-onset mitochondrial encephalomyopathy and complex I deficiency (MIM 613621);[296] sideroblastic-like anemia and iron overload (MIM 609588).[297] Most recently, *NFU1* and *BOLA3* mutations were reported in neonates with defects in lipoic acid biosynthesis and deficiency of OXPHOS complexes I, II and III (MIM 605711 and 614299).[298-300] Therefore, understanding the exact mechanism through which these proteins work, individually and in cooperation with each other, is essential not only for our basic comprehension of such biological machinery, but also for understanding how mutations in the genes coding for human homologs cause serious diseases. Over the last years, a plenty of genomic[301] and biochemical ([101] and references therein) data has been collected and several experimental structures of ISC system proteins, alone or in complex with each other, have been solved, through solution NMR and X-ray crystallography.[302-319] These studies have unravelled several structural and mechanistic aspects of the [Fe-S] cluster formation process, in general, and of how proteins belonging to the ISC system carry out their own specific tasks, in particular.

The bacterial proteins IscS and IscU are two essential components of the ISC system and, owing to their critical importance in this metabolic pathway, are highly conserved from bacteria to primates, with the eukaryotic orthologues being known as Nfs1 and Isu respectively.[320-322]

IscS[323] is a widely distributed, α-family pyridoxal-5'-phosphate(PLP)-dependent desulphurase enzyme that catalyzes the reductive elimination of sulphur from L-cysteine to yield L-alanine and an enzyme bound, highly reactive persulphide intermediate species. The persulphide acts as a source of inorganic sulphur for the assembly of the cluster, which occurs onto a transient scaffold protein. However, it can also deliver sulphur to several other sulphur-accepting proteins, which in turn commit it to a number of different metabolic pathways. Indeed, IscS is not only involved in the [Fe-S] cluster formation but it is also able to provide sulphur for e.g. molybdenum cofactor (Moco) biosynthesis and tRNA thiolation.[324-326] From a structural viewpoint, IscS exists as a symmetric, obligate homodimer, with each monomer featuring an α+β fold, with a greater fractions of residues in α helices (40.1%) than in β strand conformation (13.8%). The active site consists in a highly polar pocket located near the surface of the protein and separated by about 30 Å from that of the other monomer, on the opposite edge of the dimer. The active site pocket contains a PLP cofactor, covalently anchored by formation of an internal aldimine Schiff base with K206 and also solidly held by a number of polar and non-polar interactions, and a catalytic cysteine (C328) within a flexible loop. Throughout the production of [Fe-S] clusters, the latter, which spans residues 323-337, is thought to shuttle from the active site of IscS to that of IscU, so as to supply IscU with the inorganic sulphur that, together with iron, will be used to assemble the cluster.

IscU is the primary scaffold protein[327] and acts by providing the active site for [Fe-S] clusters assembly, transiently hosting the assembled [Fe-S] cluster and eventually delivering it to the final acceptor proteins. [Fe-S] clusters are ligated by three conserved cysteines (C37, C63, C106) and one histidine (H105) located in or near loops at one end of the IscU structure. However, not all IscU homologs from other organisms feature a conserved histidine,[328] which could then be substituted by another nearby side chain or, perhaps, even a water molecule as the fourth ligand. From a structural point of view, IscU exists in equilibrium between at least two slowly interconverting conformational states in the test tube:[329] one compactly folded (termed state S) and the other partially unfolded (termed state D). Indeed the free form of IscU exhibits molten globule-like characteristics,

lacking of a well-definite tertiary structure, with the N-terminal segment detached by the rest of the structure and highly flexible in solution. The core of the structure instead features an α-β sandwich architecture, with four α-helices packed against a three-stranded antiparallel β-sheet. Free IscU is also intrinsically unstable, being able to undergo both cold and heat denaturation at detectable temperatures,[330] though it is known that the presence of a Zn^{2+} ion or of the [Fe-S] cluster strongly stabilizes this protein. Indeed, when bound to the latter, as well as when in complex with IscS, the IscU conformational equilibrium results shifted towards its folded state (S), both within crystal structures and in solution.[302,307] It cannot be excluded that crowding effects or interactions with other partners may also stabilize IscU to its compact form in the cellular environment. However, since usually IscU is in a metal bound state, the physiological importance of the unbound state remains unproven.

During the [Fe-S] cluster assembly stage, IscS and IscU form a tetrameric S-shaped complex[307] in which the roughly prolate ellipsoidal IscS dimer interacts with two copies of IscU, one for each monomer of IscS (Figure 7.1A). In the IscS-IscU complex each copy of IscU sits close but not in direct contact with the catalytic site of IscS and binds the IscS dimer at the opposite extremes of the two poles, faraway from each other, forming a complex 150 Å long and 65 Å wide.

In-vitro the IscS-IscU complex alone was shown to be sufficient to assemble [Fe-S] clusters without the need of any further protein, provided that the reagents (cysteine, iron and a reducing agent) are present in the test tube.[327] This is not what happens *in-vivo*, where the [Fe-S] cluster assembly requires the participation of other proteins.[331] These supplementary proteins, whose tasks range from acting as electron donors or source of iron to regulating the kinetics of cluster assembly, compete for the same binding surface, represented by a cavity delimited by the IscS dimer and the IscU binding site. This leads to the formation of different ternary complexes of which the IscS-IscU binary complex constitutes the core unit. A particularly interesting trimeric complex is that with CyaY (Figure 7.1B),[332,333] since this protein is the bacterial ortholog of frataxin (FXN) which, in humans, is linked to Friedreich's ataxia, a neurodegenerative disease caused by reduced levels of frataxin.[292] Previous studies have shown that frataxin intervenes in regulating the kinetics of [Fe-S] cluster formation, although, paradoxically, in prokaryotes appears to act as an inhibitor[332-335] while in eukarotes is an activator.[336-341]

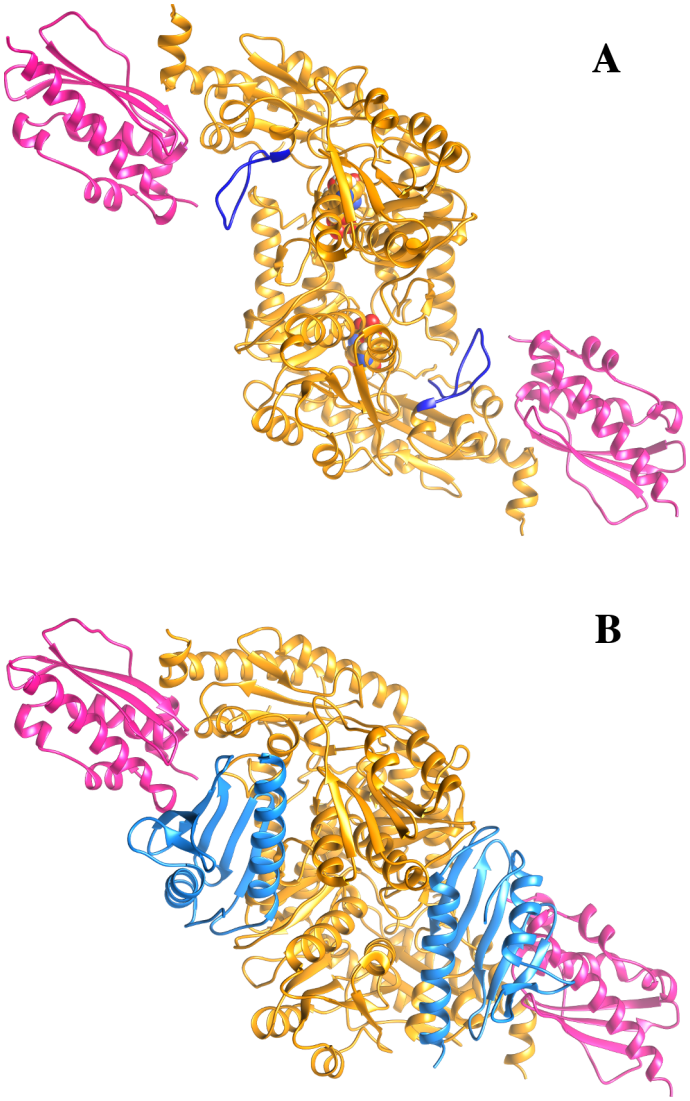


Figure 7.1: Topology of the IscS-IscU and IscS-IscU-CyaY complexes (A) IscS-IscU complex structure, formed by the IscS dimer (orange) and two IscU monomers (pink). The catalytic loop on IscS is colored in blue and the PLP cofactor is displayed as spheres. (B) IscS-IscU-CyaY complex structure, where two CyaY monomers (blue) are inserted in the cleft formed between IscS and IscU. All structures are depicted as cartoons.

Given the relevance of the IscS-IscU complex to both *in-vitro* and *in-vivo* cluster assembly, it is not surprising that a great effort has been directed towards understanding how IscS and IscU cooperate together and how such cooperation is

affected by the presence of other key players of the ISC system. Nonetheless, many steps of the biosynthesis are still not completely defined and several questions remain open about the way the IscS-IscU complex operates. It is still not clear, for instance, how the IscS flexible loop bearing C328 shuttles between the IscS catalytic site and the cluster assembly site on IscU. In one of the crystal structures of the complex (PDB code: 3LVL),[307] the catalytic loop is so flexible not to be observable. This is not the case for the structure of the IscS-IscU complex from *A. fulgidus* (PDB code: 4EB5) where the loop is close to IscU providing a potential alternative ligand for the cluster.[308]. How the dynamical behaviour of IscU and its intrinsic flexibility reflect onto complex formation with IscS is another important aspect which demands further elucidation, as well as how iron reaches the catalytic site on IscU. Finally, it would be interesting to understand how the aforementioned steps or the dynamical features of the IscS-IscU complex, e.g. the shuttling motion of the catalytic loop, are influenced by the presence of CyaY. Only a low resolution model of the ternary IscS-IscU-CyaY complex is available[333] hampering direct evaluation of how its presence close to the catalytic loop might affect it.

In order to address these open questions and have a more dynamical picture of the IscS-IscU complex than what we can gather from the X-ray structures, we have carried out an extensive computational study, the first one, to the best of our knowledge, on a complex between ISC proteins, based on overall almost 1μs molecular dynamics simulations. Specifically, we have simulated two systems: I) a binary complex, made up by IscS and IscU only, and II) a ternary complex, made up by IscS, IscU and CyaY. Our results identify motions taking place at the binding interface between IscS and IscU which are potentially connected to the biological activity of the complex. We also show how the presence of CyaY stabilizes the IscS-IscU interaction and provide evidences for a possible mechanism explaining how CyaY reduces the kinetics of the cluster assembly process. Our data bear important consequences for understanding the role of CyaY in [Fe-S] protein biogenesis.

7.1 Computational Methods

7.1.1 Systems preparation

IscS-IscU binary complex. The *E. coli* IscS-IscU complex (PDB code: 3LVL)[307] was used for molecular dynamics study of the binary complex. The missing segment of the IscS loop bearing the catalytic cysteine (residues 328-334) was reconstructed using the Modeller[189] module (v9.12) as implemented into the UCSF-Chimera software.[197] Five different conformations of the loop were generated. In order not to artificially orient the dynamics of the loop towards one side or another of the binding interface, we selected as starting conformation the one where the loop was as much as possible equidistant from both IscS and IscU. The same coordinates were applied to both missing loop segments, so as to avoid possible inconsistencies. The complex was then solvated in a ~ 23 Å layer cubic water box using the TIP3P water model parameters. Na^+ and Cl^- ions were added so as to ensure system electroneutrality and to set the final concentration to 0.15 M. The final system size was 188 Å x 95 Å x 105 Å for a total number of atoms of ~170000. The system was minimized in two stages: first, a 20000-step run was carried out with restraints on all the protein atoms (5 kcal/mol/$Å^2$); then, a further 10000-step minimization was carried out by applying restraints on the cofactor and C_α protein atoms only. Then, a small (200 ps) NPT simulation at 200 K and 1 atm was performed with restraints on all the protein atoms (5 kcal/mol/$Å^2$), in order to adjust the volume of the simulation box, while preserving the minimized protein structure obtained in the previous steps. Afterwards, the system was slowly heated up to 300 over a 3 ns period, gradually releasing the restraints to 1 kcal/mol/$Å^2$ along the thermalization process. Subsequently, the system was equilibrated for 2 ns, gradually reducing the restraints to zero.

IscS-IscU-CyaY ternary complex. The *E. coli* IscS-IscU-CyaY complex structure[333] was used for molecular dynamics study of the ternary complex, by gracious permission of Pastore *et al*.. The procedure to prepare the system was slightly different, in order to minimize the steric clashes between the CyaY monomers and the IscS-IscU subunits. First, the side chains of the residues at the interface between the CyaY monomers and the corresponding IscS-IscU subunits were manually adjusted by means of the UCSF-Chimera software package.[197] Then, the simulation box was prepared in the same way as did for the IscS-IscU

complex. The final system size was 194 Å x 102 Å x 111 Å for a total number of atoms of ~205000. The system was then minimized and equilibrated in the following fashion. A first round of 15000-step energy minimization was performed with restraints on the C_α protein atoms and on the cofactor atoms (5 kcal/mol/Å^2), followed by a subsequent 20000-step energy minimization run with no restraints, so as to minimize as much as possible steric clashes in the first stage. Then, a small (200 ps) NPT simulation at 200K and 1 atm has been performed with restraints on all the protein atoms (5 kcal/mol/Å^2), in order to adjust the volume of the simulation box, while preserving the minimized protein structure obtained in the previous steps. Afterwards, the system was slowly heated up to 300 K over a 3 ns period, gradually releasing the restraints to 1 kcal/mol/Å^2 along the thermalization process. Finally, the system was equilibrated for 6 ns, gradually reducing the restraints to zero.

7.1.2 Simulation and analysis details

Molecular dynamics. Production runs were performed under NPT conditions at 1 atm and 300 K and extended up to 400ns for both complexes. A 10 Å cutoff (switched at 8.0 Å) was used for atom pair interactions. The long-range electrostatic interactions were computed by means of the PME method using a 1.0 Å grid spacing under PBC. The RATTLE algorithm[342] was applied to constrain bonds involving hydrogen atoms, and thus an integration 2 fs time step interval could be used. All the systems were simulated with NAMD[190] (v2.9), using the ff99SBildn Amber force field parameters[343,344] for protein and ions. The parameters for the PLP were generated in two steps. Initially, charges were computed using the restrained electrostatic potential (RESP) fitting procedure.[345] The ESP was first calculated by means of the Gaussian09 package[247] using a B3LYP/6-31G* level of theory, and then the RESP charges were obtained by a two-stages fitting procedure using the program R.E.D..[248,346] Missing bonds, angles, torsion and improper torsion angle parameters were then generated using Antechamber.[347]

Trajectory analyses. Analyses were performed using the cpptraj[131] tool and in-house scripts exploiting the MDAnalysis library.[196] The Principal Component Analysis has been performed by means of the Gromacs analysis tools[348] g_covar and g_anaeig, restricting the analysis to the C_α atoms only. The

electrostatic potential was evaluated using the APBS server.[134] Plots were created using the Matplotlib software.[198] Figures were generated using the UCSF-Chimera software package and the VMD program.[188]

7.2 Results

7.2.1 Stability and structural fluctuations of the IscS-IscU binary complex

First, we evaluated the system RMSD, in order to assess its stability throughout the simulated time interval. (Figure 7.2) After the first 60 ns the complex reached structural relaxation, as indicated by the RMSD plateau. Thus only the portion of the trajectory from this time onward was employed for further analyses.

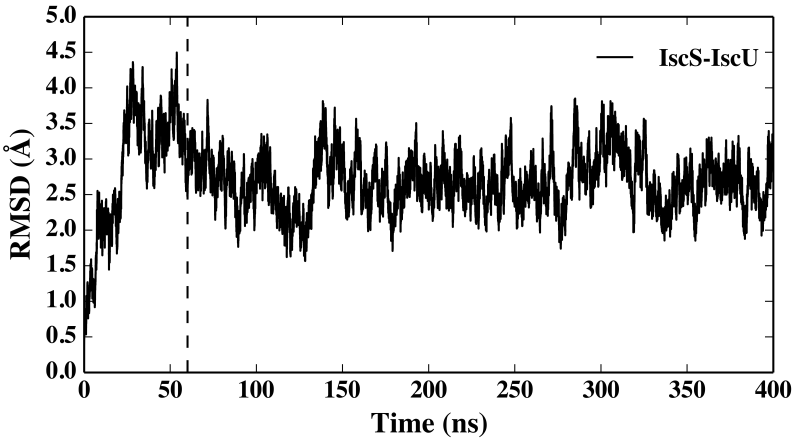


Figure 7.2: **Structural deviations of the binary system as a function of time.** Backbone RMSD of the binary system with respect to the starting frame. At the plateau (~60 ns) the RMSD fluctuates around an average value of about 2.6 Å.

Throughout the trajectory, the overall structures of IscS and IscU remained mostly unchanged, with no major unfolding or openings of IscU or IscS. The N-terminal portion of IscU, which in the apo-structures was shown to be rather disordered,[302,304] remained quite rigid during the simulation, adopting a helical

conformation, as observed in all the crystal structures of the complex. Besides, IscU remained solidly bound through the interface that involves residues I8, Y11, E12, V40, K42, Y61, I67 and K103 of IscU and residues E309, E311, S312, M315, E347, R379, L380, P385, L386 and E388 of IscS. Specifically K42, Y61 and K103 on IscU engaged in salt bridges or hydrogen bonds E388, E347 and E311 on IscS, respectively, whereas the remaining residues were mostly involved in non-polar interactions between each other and in other non-specific contacts with the backbone atoms of either IscS or IscU (Figure 7.3). However, despite such a firm anchoring way, we noted that IscU gently rocked with respect to IscS around a pivot formed by their interaction surface.

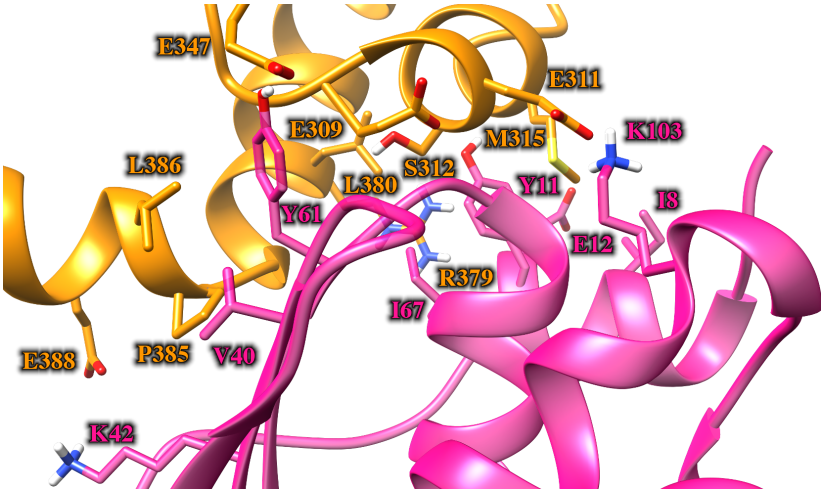


Figure 7.3: **IscS-IscU complex interface.** Zoomed view of the interfacial region between IscS (orange) and IscU (pink). The side chains of the main interacting residues are shown as sticks, while the rest of the protein is depicted as cartoons.

Analysis of the structural fluctuations of the complex through the evaluation of the RMSF revealed three main flexible regions of the complex (Figure 7.4). The first one is represented by the C-termini of the two IscS protomers. The second one is a small β-turn (residues 34-39) on IscU, bearing one of the three conserved cysteines (C37) involved in the coordination of the cluster. Interestingly, we observed an opening motion of such β-turn in one of the two IscU monomers, leading to an increased exposure of the active site of IscU (Figure 7.5). The last flexible region is represented by the catalytic loop on IscS, in agreement with the inherent flexibility such a loop should exhibit in order to move between two active

sites distant more than 15 Å.[307] For this reason, the dynamics of this loop was examined in more detail, as described in the next paragraph.

Moreover, the RMSF analysis showed relatively high structural fluctuations distributed among several secondary structure elements of IscU, particularly α-helices but also a few β-turns and β-sheets (comprising residues ca. 21-28, 41-56, 64-87 and 108-125).

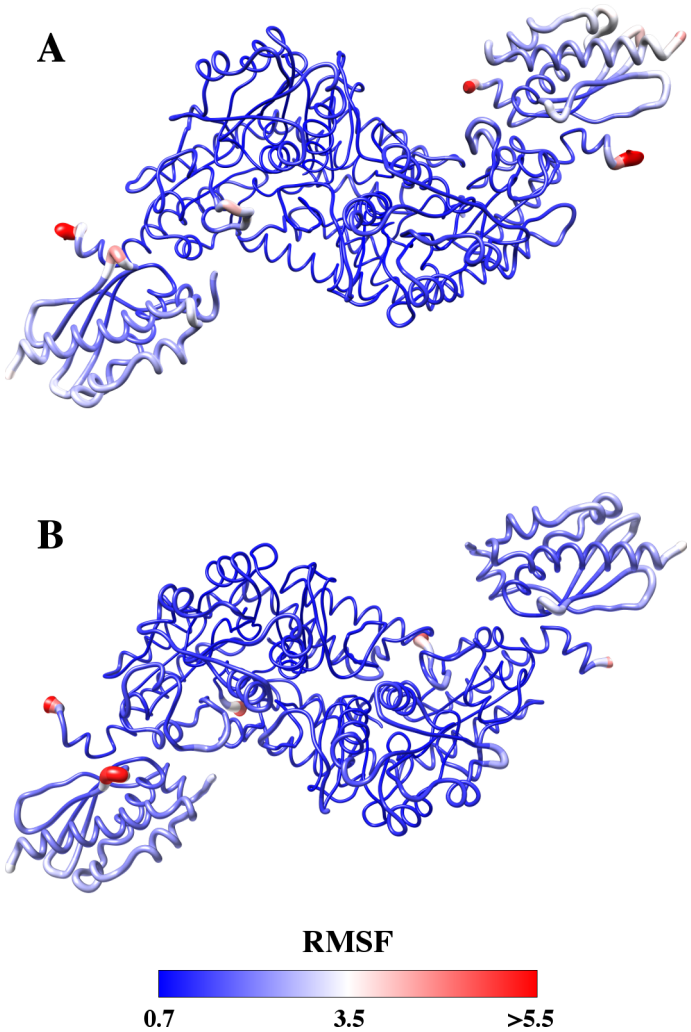


Figure 7.4: Backbone RMSF. Average structure of IscS-IscU in (A) binary and (B) ternary complex, depicted as ribbon and colored according to backbone RMSF values, from lowest (blue) to highest (red). The thickness of the ribbons is proportional to RMSF values.

Intriguingly, these residues are located mainly in the tip of IscU that is fully exposed to the external medium, as opposite to that at the interface with IscS, which instead appeared rather stable.

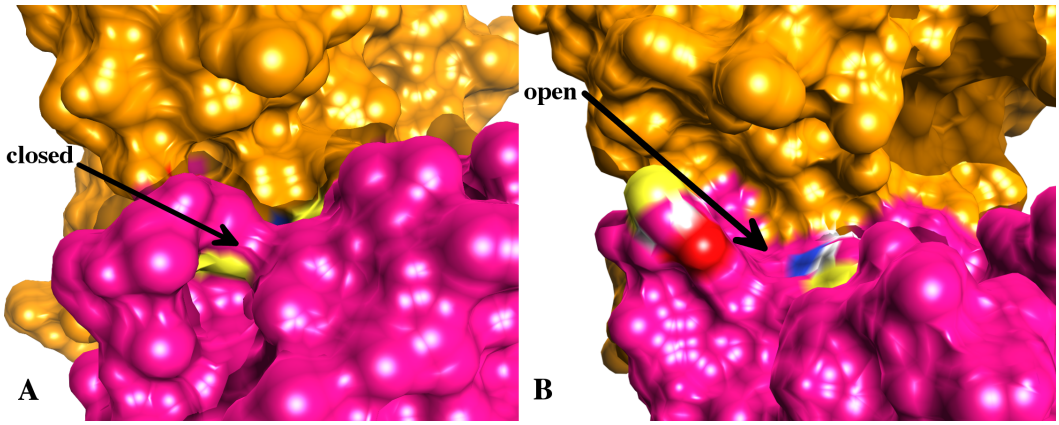


Figure 7.5: Opening motion of the IscU β-turn. Differences in the IscU active site accessibilities between (A) the IscS-IscU average structure, as issuing from our MD simulation, and (B) the E. coli IscS-IscU crystal structure (PDB: 3LVL). In both panels, IscS and IscU are displayed as surfaces, colored in orange and pink, respectively.

7.2.2 Dynamics of the IscS catalytic loop

Owing to its relevance in the catalytic process, we analyzed in more detail the dynamics of the IscS catalytic loop, which is responsible for sulphur transfer to IscU. We monitored three different parameters in order to identify structural changes possibly correlated with the transferring mechanism: I) the secondary structure evolution; II) the gyration radius and III) the distance between the C328 sulphur atom and those belonging to C63 and C106 in the IscU active site. In the latter case, due to the observed flexibility of the β-turn carrying C37, the sulphur atom of C37 was not considered in the analysis so as not introduce an artificial bias in the distance calculation.

Intriguingly, our simulation showed a conformational transition in one of the two catalytic loops in the IscS-IscU complex (Figure 7.6). After about 125 ns, a small portion of the loop, namely residues C328 to S332, underwent a structural rearrangement from a mostly 3_{10} helical structure to a conformational state

characterized by an equilibrium between a β-turn and a 3_{10} helical structure, which was maintained throughout the rest of the simulation (Figure 7.6A).

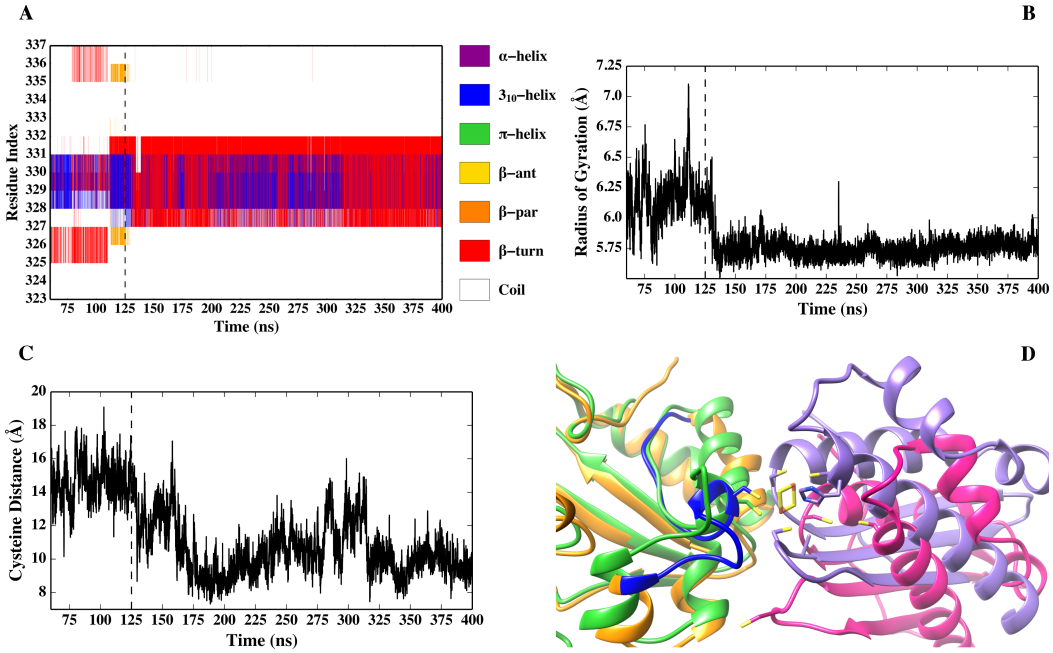


Figure 7.6: Conformational transition of the IscS catalytic loop. Time evolution of the structural rearrangement of the IscS catalytic loop as monitored by (A) secondary structure, (B) radius of gyration (backbone atoms only); (C) distance between C328 and the IscU active site (C63 and C106). (D) Structural superposition of a representative configuration of the IscS(orange)-IscU(pink) complex (the catalytic loop is in blue) taken from the last part of the MD simulation and the *A. fulgidus* IscS(green)-IscU(purple) crystal structure. The superposition was carried out by fitting on the IscS dimer only. C328 on the IscS catalytic loop and the conserved cysteines on IscU are shown as sticks, as well as the [Fe-S] cluster bound in the *A. fulgidus* structure. Note how the structural transition occurring at about 125 ns in our simulation (panel A-B), which triggers a progressive approach of C328 towards the active site of IscU (panel C), brings the catalytic loop in a conformation very similar to that observed in the *A. fulgidus* crystal structure (panel D). Despite that, the IscU active site remains far apart from the catalytic loop, as compared to the *A. fulgidus* structure.

Almost simultaneously, the radius of gyration of the loop decreased, suggesting an increase in the "compactness" of the loop (Figure 7.6B). This conformational transition, which is likely triggered by the breaking of an adjacent β-turn (residues 325-327), enabled a progressive approach of C328 towards the active site of IscU up to a final distance of ~9 Å (Figure 7.6C). Such distance is relatively close to that observed in the X-ray structure of the *A. fulgidus* IscS-IscU complex (~6 Å), in which the cluster is completely formed and coordinated by the catalytic loop C328. Indeed, structural superposition of the IscS loop from our simulation and the holo crystal structure showed that the two loop conformations are quite similar (Figure 7.6D). On the contrary, the structures of the IscU monomers at the interface with IscS differ significantly in the two complexes. Specifically, IscU appeared somewhat closer to IscS in *A. fulgidus* crystal structure than in the simulated apo-complex.

Accordingly, when we measured the geometrical angle formed by the intersection between the axis connecting the centre of mass of the two IscS monomers and the axis of the IscU α-helix spanning residues ~100-120 (Figure 7.7) we noted a significant difference between the two complexes, i.e. ~142° and ~125° respectively. These observations suggest that, in addition to the movement of the IscS catalytic loop, other structural transitions could be required within the complex in order to bring C328 at contact distance with the IscU active site.

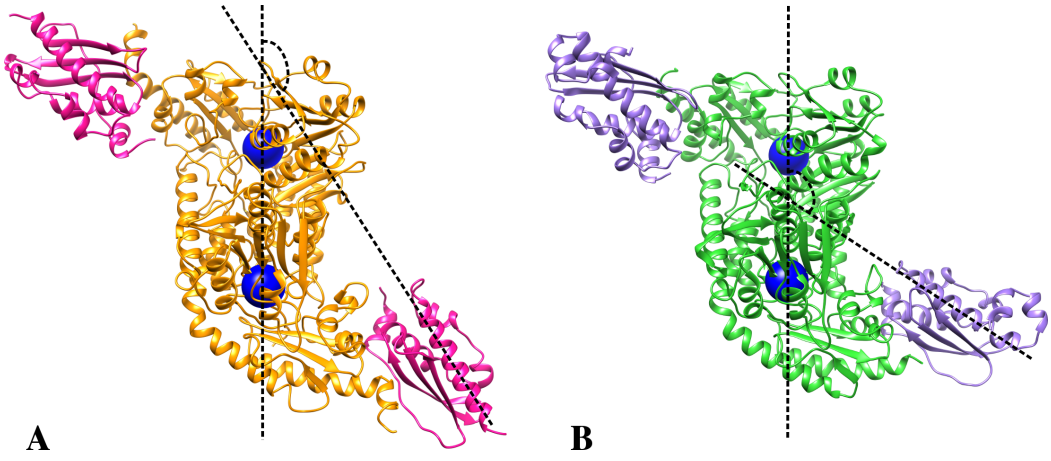


Figure 7.7: IscU orientation within the binary complex. Angle formed between IscU and IscS from (A) the average structure issuing from our simulation and (B) the *A. fulgidus* complex crystal structure. Both structures are depicted as cartoons.

Therefore, with the aim of identifying other dynamical features of IscS or IscU potentially involved in the sulphur transfer process, we analyzed in more detail the global correlated motions of the complex.

7.2.3 Collective motions analysis of the IscS-IscU complex and interface dynamics

Prompted by the observation of the IscU rocking motion with respect to IscS and by the structural superposition reported in the previous paragraph, we performed a PCA of the binary complex dynamics, aiming at identifying the most significant collective motions. The analysis showed that the highest amplitude motions of the complex are mainly located in correspondence of the IscS catalytic loops, the IscU β-turns and, more in general, the whole IscU monomers, as illustrated in the first eigenvector (which alone accounts for >50% of the total variance of the system) (Figure 7.8).

These results are in good agreement with the outcome of the RMSF analysis. Furthermore, a closer inspection of the collective motion described by the first eigenvector revealed that the IscU dynamics was characterized by a large amplitude, coherent motion, resembling a sort of "rocking" of the IscU monomers as a whole around their respective interfaces with the IscS dimer. Similar indications emerged also from the analysis of the second and third eigenvectors (data not shown). To further support the soundness of the rocking motion suggested by the PCA results, we evaluated the range spanned by the geometrical angle between IscU and the IscS dimer, (as defined in the previous paragraph) and it was rather wide (~123° - ~161°), as illustrated in Figure 7.9A.

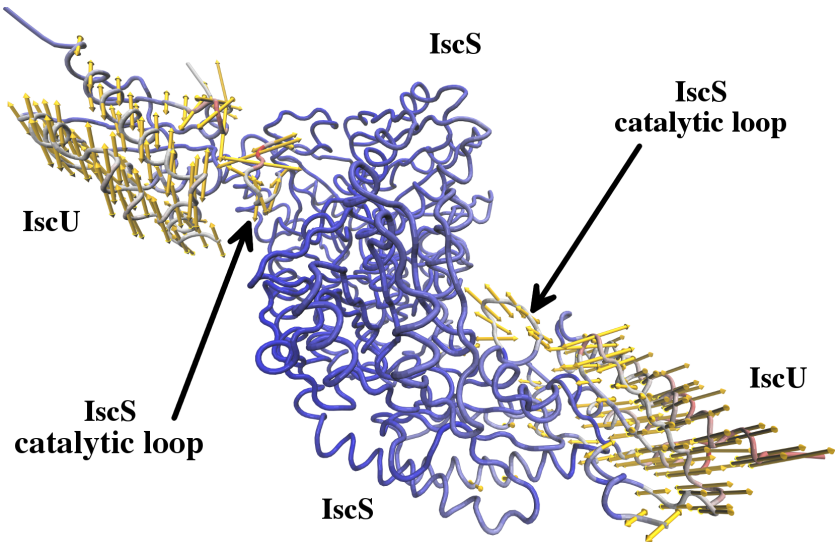


Figure 7.8: Main collective motion of the IscS-IscU complex. Pictorial representation of the main collective motion of the IscS-IscU complex, as issuing from the PCA (Cα atoms only), showing the IscU pivotal movement around the IscS dimer and the motion of the IscS catalytic loop. The contribution of each residue to the motion along the first PCA eigenvector, is mapped onto the IscS-IscU complex structure (depicted as ribbons), from lowest (blue) to highest (red). Vectors display the amount (i.e. length) and direction of the residue motion along the eigenvector. Note that the longest arrows are concentrated on the IscS catalytic loops and on IscU protomers, especially on the external tips. For the sake of clarity, vectors on residues bearing only a minor contribution to the motion along the eigenvector (i.e. whose length is lower than 2Å) where omitted.

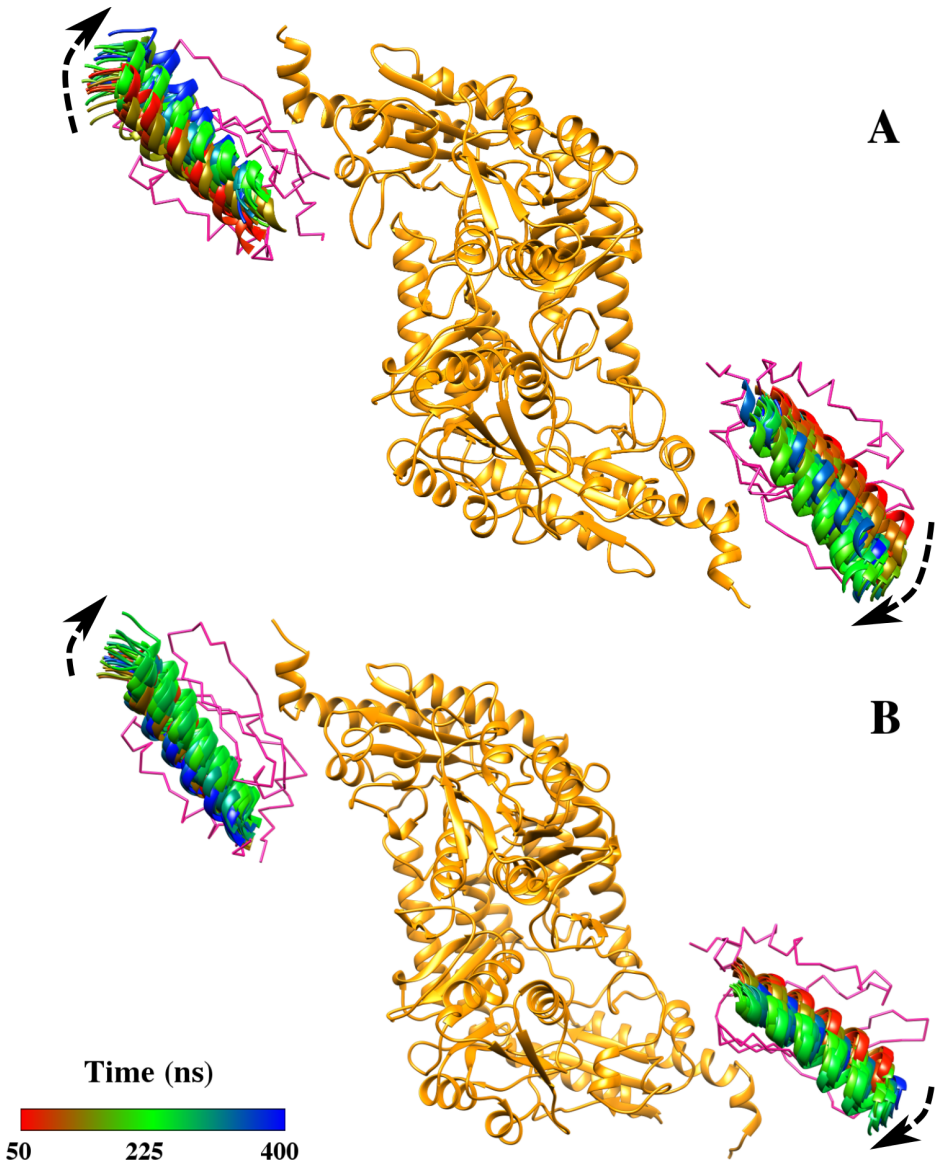


Figure 7.9: Dynamics of the IscU monomers at the interface with IscS. A set of uncorrelated frames (about 1 every 10 ns) was extracted from the simulation (A) binary and (B) ternary complexes and fitted onto the IscS dimer structure, depicted as orange cartoons. These frames represent different orientations of the IscU α-helix that was used to define the IscU angle with respect to IscS, each depicted as cartoons and colored as a function of simulation time, while the rest of IscU is displayed as pink ribbons.

Finally, we performed an analysis of the residue-wise dynamic cross correlations (DCC). The DCC map in Figure 7.10 emphasizes the significant correlations (R≥0.5), colored according to their relative intensities. A careful inspection of the map evidenced several positive correlations among the residues belonging to the same IscU monomer. In particular, most of the highest off-diagonal (i.e. non-self) correlations (R≥0.75, colored in red and brown) clustered among the secondary structure elements already identified as the most flexible regions, plus the external tip of IscU, which is not involved in IscS binding. Altogether these findings indicate that the binding of IscU at the interface with IscS is not very tight and that the IscU monomers, despite the stability of their secondary and tertiary structures, retain a certain degree of residual mobility upon formation of the complex.

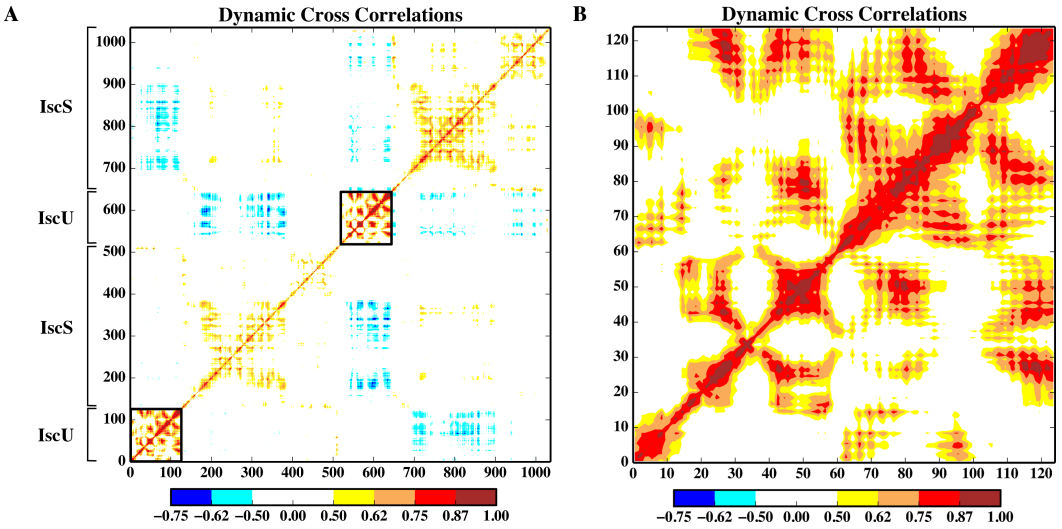


Figure 7.10: DCC analysis. (A) The dynamic cross correlation map of the binary complex shows that the highest correlations (>0.75) are concentrated among the single IscU monomers. (B) Magnified view of the correlation map of one of the two IscU protomers.

7.2.4 Stability of the IscS-IscU-CyaY ternary complex and role of electrostatics in CyaY recognition

We investigated the structural and dynamical features of the IscS-IscU-CyaY ternary complex and compared them to those of the binary system. Since no high resolution structure of the ternary complex still exists, we used as the starting structure a recently proposed low resolution model obtained through an experimentally-driven protein-protein docking procedure.[333] After a careful equilibration (see Methods for details), the IscS-IscU-CyaY complex showed a high stability in the time scale of our simulation, as indicated by the RMSD analysis (Figure 7.11), in spite of the low-resolution of our starting model. We then used the optimised model to analyze aspects of the trimeric complex internal dynamics, possibly related to the CyaY functionality. As in the previous case, we restricted all our analyses to the equilibrated portion of the trajectory only.

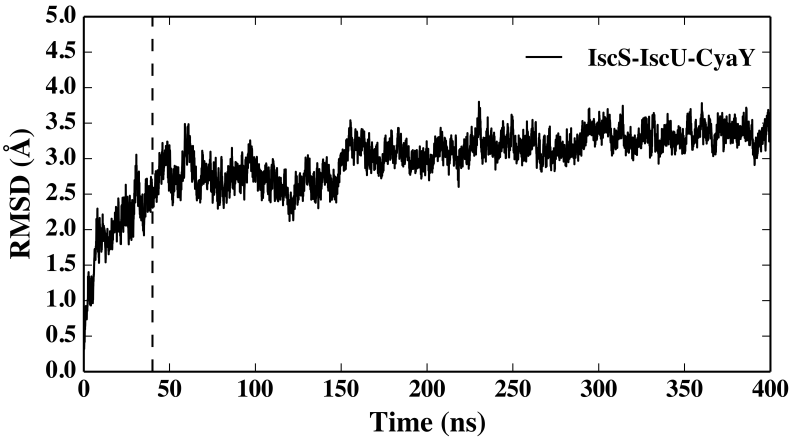


Figure 7.11: **Structural deviations of the ternary system as a function of time.** Backbone RMSD of the ternary system with respect to the starting frame. At the plateau (~40 ns) the RMSD fluctuates around an average value of about 3 Å.

The ternary complex features the partial insertion of CyaY in the cleft delimited by the IscS dimer interface, on one side, and the IscU-binding site, on the other side (Figure 7.1B). Within the complex CyaY packs mainly against IscS, far from the PLP site (~18Å) but in proximity of the catalytic loop, which is contacted at residues A327, T329, S330, S332 and E334 via a few residues located on CyaY β-sheets (namely, N35, V38, N52, R53). From the very beginning, the

nature of the binding interface appears quite complex. CyaY weakly contacts IscU through the surface of its β-sheets establishing almost exclusively hydrophobic interactions. In particular, the exposed residue W61 of CyaY is close to the conserved cysteine C37 of IscU and to a few other residues, including P35, A36, P101, V102 and I104. Conversely, the picture of the interface with IscS appears diametrically opposite, being mainly characterized by the direct recognition between a positively charged patch on IscS and a negative region on CyaY, as well as several other polar and non-polar contacts, as (Figure 7.12) shows.

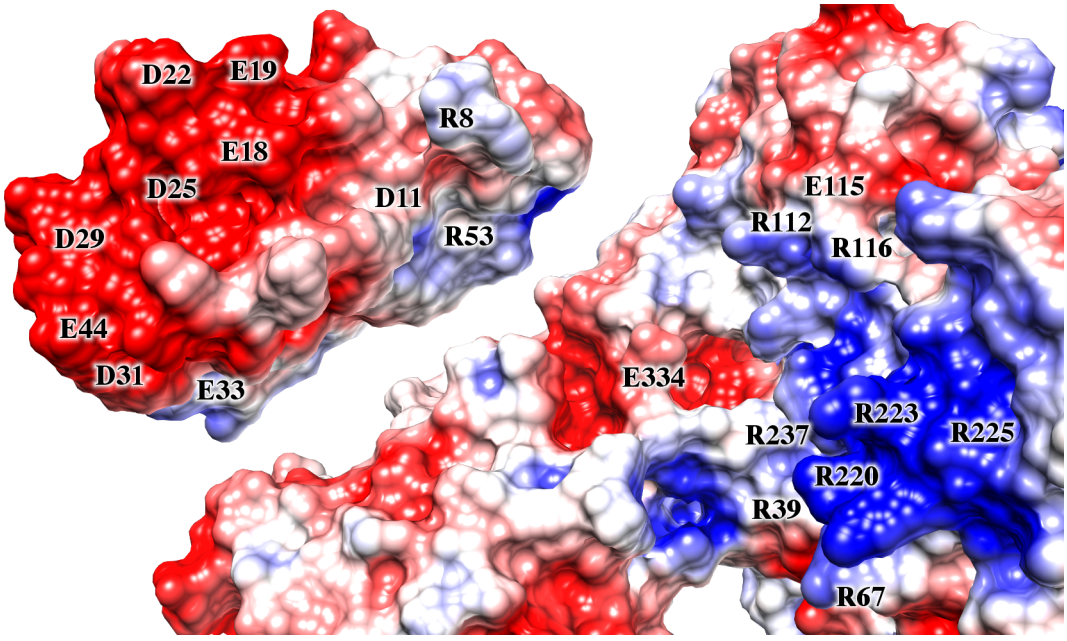


Figure 7.12: Electrostatic potential surfaces for the IscS-IscU complex (right) and CyaY(left).

Such patch on IscS represents a binding site that presumably also hosts several other IscS partners, amongst which there are TusA[307] and the ancillary proteins Ferredoxin[349] and IscX,[350] being, in all three cases, complementary to the mainly negative charge of these partners. This indicates that the interaction with IscS is primarily electrostatic in nature. Then, in order to identify, in a dynamical framework, the residues potentially more important for the binding, we performed an analysis of the principal salt bridges established at the interface between CyaY and IscS. The results are summarized in (Table 7.1). Surprisingly, we found a relatively higher number of interacting residue pairs than we expected based on

previous literature reports.[307,332,333,351,352] These pairs involve residues R8, D11, E18, E19, D22 D25, D29, D31, E33, E44 and R53 on CyaY and R39, R67, R112, E115, R116, R220, R223, R225, R237 and E334 on IscS and are depicted in (Figure 7.13). Interestingly, positively and negatively charged residues were present on both CyaY and IscS surfaces.

CyaY	IscS	% Persistence
D11	R112	36.2
D22	R225	92.1
D25	R223	61.5
D25	R220	92.1
D29	R67	91.3
D29	R237	78.9
D29	R220	78.3
D31	R67	85.6
E18	R112	99.0
E19	R116	84.7
E33	R39	97.0
E44	R67	54.8
R53	E334	99.2
R8	E115	75.6

Table 7.1: Interface salt bridges established within the ternary complex. All charged residues (Arg, Lys, Glu, Asp) whose side-chain nitrogen-oxygen distance was below 5 Å in at least one frame were considered for the analysis. Only the salt bridges persisting for a significant fraction of the total simulation time (>35%) are reported, along with the highest percentage observed in both monomers. The analysis has not revealed any salt bridge between IscU and CyaY.

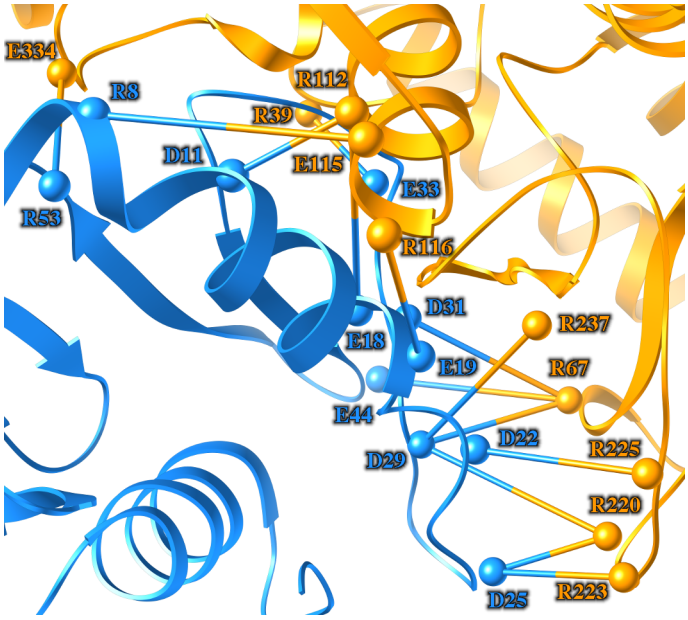


Figure 7.13: IscS-CyaY interface. Zoomed view of the interface between the IscS dimer (orange) and CyaY (blue). The more persistent salt bridges established with CyaY are depicted as sticks connecting the Cα atoms (spheres) of the corresponding residues, as evidenced by the salt bridge analysis (see Table 7.1).

7.2.5 The effect of CyaY on the IscS-IscU dynamics

After the nature of the binding interfaces, we also characterized in more detail the dynamics of the IscS-IscU-CyaY complex. Analysis of the fluctuations revealed a significant reduction of the structural flexibility of the IscS-IscU system upon CyaY binding (Figure 7.4B), the only residues displaying notable fluctuations being located on the IscS C- and N-termini and on the IscU β-turn mentioned earlier. The catalytic loop on IscS was considerably more rigid and didn't show any relevant motion towards the IscU active site, in contrast with what we previously observed in the binary complex. The external tip of IscU retained some degree of flexibility, though it was found to be less flexible in the ternary complex (Figure 7.9B), suggesting that the presence of CyaY strengthens the binding between IscS and IscU. We also tried to perform a DCC analysis but comparison of the DCC maps for binary and ternary systems didn't evidence significant differences (data not shown).

To further investigate other potential differences between the binary and ternary complexes, we performed an ED analysis on both systems, restricting it to the IscS and IscU monomers only, for the sake of direct comparison. Specifically, we first performed a PCA on the ternary system and then projected the MD trajectories of binary and ternary systems onto their respective first two eigenvectors, which in both cases account together for >50% of the total variance. A comparison of the 2D essential subspaces exploration (Figure 7.14) highlights two main differences between the systems considered.

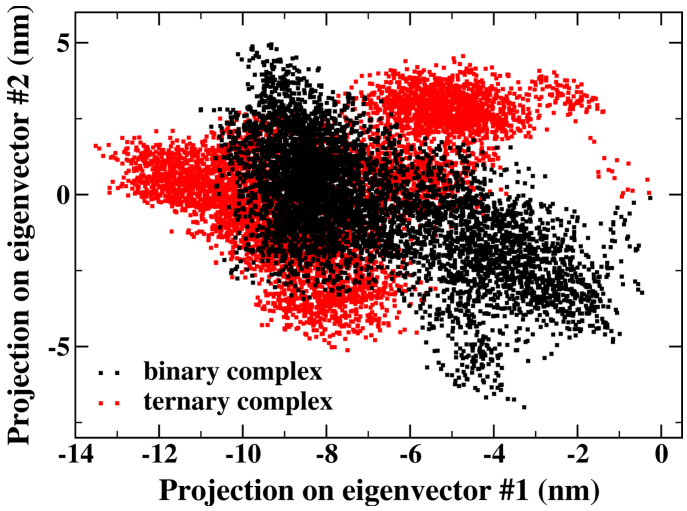


Figure 7.14: Essential dynamics analysis of the binary and ternary complexes (1). Projection of the simulated trajectories of the IscS-IscU tetramer (Cα atoms only) for both systems onto the 2D ED plane described by their respective first two eigenvectors.

Firstly, the ternary complex covers a slightly smaller conformational subspace with respect to the binary system, despite its slightly larger conformational sampling (~20ns), remaining closer to its respective average. This further confirms that the presence CyaY can indeed play a role in hampering, at least in part, the overall IscS-IscU dynamics. Moreover, the essential plane regions explored by the two systems do not superimpose much. This may be due either to an effective difference in the sampling of the same collective motions in the two cases or to a more general and intrinsic difference between the two essential spaces. The small normalized overlap (~0.261) between the eigenvector sets of the two systems

indicates that the large amplitude motions of the two systems are, altogether, quite different. Therefore, we projected the trajectory of the ternary complex onto the eigenvectors of the binary complex, so as to verify whether the collective dynamics of the IscS-IscU system in the former actually resembles that of the CyaY-free system, e.g. in the motion of the IscS catalytic loop or in the IscU rocking motion. Figure 7.15 shows a comparison between such projection and that of the binary complex trajectory onto its own eigenvectors. In this case, the conformational subspace exploration of the ternary system is even more limited, being the dynamics confined within only one "basin" of the essential plane, and not at all superimposable with that of the binary system.

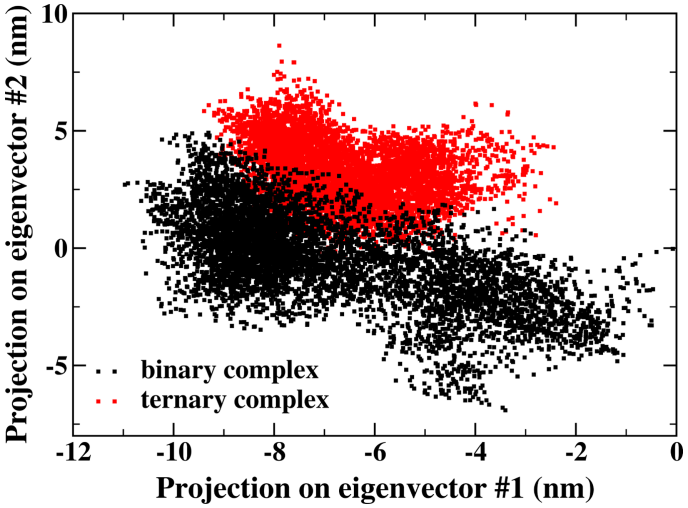


Figure 7.15: Essential dynamics analysis of the binary and ternary complexes (2). Projection of the simulated trajectories of the IscS-IscU tetramer (Cα atoms only) for both systems onto the 2D ED plane described by the first two eigenvectors of the binary complex. The analysis highlights the smaller conformational subspace explored by the IscS-IscU tetramer within the ternary complex, as compared to the binary system.

When we analyzed the per-residue contribution to the fluctuations of the two systems along their respective first eigenvectors (Figure 7.16), we noted two major differences. First, the RMSF peaks corresponding to the IscS catalytic loop in the binary system are considerably smaller in the case of the ternary system. Second,

the fluctuations along the main eigenvector of one IscU monomer experience an appreciable decrease in intensity, in the case of the latter system.

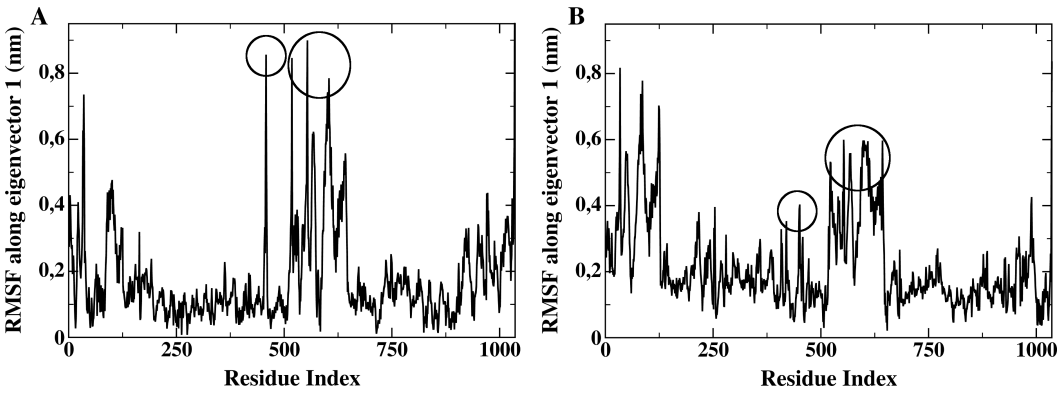


Figure 7.16: RMSF along the main eigenvector. Per-residue RMSF of the IscS-IscU tetramer in binary (A) and ternary (B) complexes along their respective first eigenvectors. Note in particular that in the ternary complex the fluctuations of one of the two IscU monomers (residues 519-643 ca) and one of the two IscS catalytic loops (residues 448-462 ca) along the first eigenvector are significantly reduced as compared to the binary complex.

These results are in agreement with our previous findings and, overall, suggest that the presence of CyaY has a noticeable impact on the dynamics of IscS-IscU, by hindering, or completely abolishing in the case of the IscS catalytic loop, the largest amplitude collective motions which characterize the dynamics of the complex in absence of CyaY.

7.3 Discussion

Atomistic MD simulations of the IscS-IscU system, both in presence and in absence of CyaY, have provided several new insights into the structural and dynamical features of such a complicated protein complex. In particular, the simulation of the binary system has identified a few relatively flexible regions of the IscS-IscU complex. We have observed the opening motion of a β-turn on IscU, which spans residues 34-39 and bears one of the three conserved cysteines involved in cluster coordination. This motion may be relevant for the activity of

the complex and the mechanism of formation of the cluster and possibly explain why absence of C37 makes the cluster coordination more unstable.[328] However, it would not be possible once the cluster is assembled, forcing the mouth formed by the three loops carrying the coordinating cysteines to stay at a fixed distance. Another flexible region identified was the IscS catalytic loop, whose flexibility is thought to reflect its biological function, i.e. shuttling between two very distant sites in order to supply the sulphur to recipient proteins. Data from our simulation suggest that such shuttling motion may be accompanied by a structural transition leading to a more compact and structured conformation. Though we found that this was sufficient to bring the loop in a conformation quite similar to that observed in the holo complex, it was not enough to bring C328 at the same distance from the IscU active site cysteines in the same structure. However, we cannot completely rule out that the differences we observed with the *A fulgidus* structure could be merely ascribed to a different crystal packing. Nonetheless, this result allows us to speculate that, set aside further small structural adjustments of the IscS loop to ensure the proper geometrical arrangement of the C328 sulphur around the cluster, an approach of IscU towards the IscS loop may also be necessary to bring C328 at coordination distance from the cluster. Intriguingly, the remarkable secondary and tertiary structure similarity shared by the IscU monomers in different functional states, i.e. bound to IscS both in absence (PDB code: 3LVL) and in presence (PDB code: 4EB5) of the cluster or as a homotrimeric complex hosting the cluster (PDB code: 2Z7E),[305] after their structural superimposition as single subunits (data not shown), seems to rule out that such an approach may arise from a internal conformational change of IscU, in favour of a "rigid-body" kind of approach. In turn, this implies a sufficiently loose binding at the IscS-IscU interface so that the IscU monomer as a whole can maintain a certain degree of overall flexibility upon formation of the complex. In this respect, the outcome of our analyses overall clearly indicate that the binding of IscU at the interface with IscS is not very tight, allowing for a certain degree of movement, and support the hypothesis of a "rigid body" motion of the IscU monomer. Note that this kind of loose binding of IscU to the IscS interface could correlate well with the relatively weak affinity of the two proteins (K_d of ~1 μM) and is also soundly consistent with the biochemical role of the complex itself. Indeed, the latter will ultimately need to disassemble to allow the transfer of the formed cluster to their final recipient proteins. Moreover, the relatively weak substrate selectivity of IscS in the sulphur transfer reaction[323,353,354] in principle implies, among other factors, a relatively not

specific, i.e. strongly complementary, binding interface, so as to recognize different partners via protein-protein interactions. Finally, a certain degree of plasticity of the complex quaternary structure may also be well-received in order to allow the accommodation of other proteins that play a role in [Fe-S] cluster synthesis.

The simulation of the ternary complex has revealed several interesting aspects of how IscS-IscU interacts with its CyaY counterpart. In particular, we have shown that the ternary complex is very stable, with CyaY solidly held to the interface with the IscS dimer via a network of salt bridges between positive and negative residues from both proteins, on one side, and to IscU via aspecific hydrophobic contacts, on the other side. Such an asymmetric nature of the binding interactions correlates with the evidence that, in the test tube, CyaY is able to form a binary complex with IscS but it is insufficient to stabilize a stand-alone CyaY-IscU complex.[333] We have thus characterized in detail the nature of the IscS and CyaY residues establishing the strongest electrostatic interactions. Besides the arginine triplet R220, R223 and R225 on IscS and the E18, E19 and D22 triplet of negative residues on CyaY, whose replacement is already known to be sufficient to completely abolish the binding between IscS and CyaY,[333] we also identified several other CyaY residues interacting at the interface, namely R8, D11, D25, D29, E33, E44, E53. It may be interesting to investigate whether point-wise or collective mutations of these residues have any effect on complex formation and/or functionality.

Our data also clearly indicate that the presence of CyaY overall dumps the fluctuations of the IscS-IscU complex. In addition to such a generalized dumping, our results seem also to indicate that the presence of CyaY may impinge on some of the functional motions identified in the simulation of the binary complex. In particular, we have observed that not only the mobility of the IscU monomers as a whole was slightly reduced, but also that the characteristic motion of IscU at the interface with IscS was more limited, as compared to the binary complex. These evidences suggest that a strengthening of the interaction of IscU with its counterpart may ensue from the insertion of CyaY at the interface between IscS and IscU, indicating a possible cooperative nature of the binding. Dumping of the fluctuations observed in our simulation can indeed be explained in the view that we know experimentally that the affinity of the IscS-IscU complex is increased by at least one order of magnitude when also CyaY is bound.[333] Moreover, we also noted that, in the time scale of our simulation, the flexibility of the IscS catalytic

loop was almost completely abolished. In this respect, the formation of a persistent salt bridge between R53 on CyaY and E334 on the IscS catalytic loop, along with other factors such as steric hindrance, may account for the observed reduced mobility of the loop in the ternary complex. It may be interesting to verify whether the replacement of these residues with alanine or residues of opposite charge can lead to an increase of the mobility of the loop.

In conclusion, despite the low resolution of our starting model of the IscS-IscU-CyaY complex, the differences detected between the binary and ternary systems allow us to propose a possible explanation for the molecular mechanism through which CyaY slows down the enzymatic rates of [Fe-S] cluster formation. Indeed, we can imagine at least two different, not mutually exclusive, ways through which CyaY may carry out its inhibitory role of the cluster assembly process. The first one involves abolishing the shuttling motion of the IscS catalytic loop and preventing the sulphur delivery to IscU, thereby blocking the process at the stage of persulphide formation on C328. The second one also affects the sulphur transfer to IscU, but via hindering the apparently necessary approach of IscU towards C328 on the IscS catalytic loop, by stabilizing the IscU binding within the complex. Both these mechanisms find additional support in the evidence that the presence of CyaY seems to be compatible neither with the arrangement of IscU nor with that of the IscS catalytic loop in the *A. fulgidus* structure, where the [Fe-S] cluster is fully formed (as shown in Figure 7.17). Indeed, the orientation of IscU with respect to the IscS dimer in the *A. fulgidus* structure, being characterized by a narrower angle between the two, would result in severe steric clashes with CyaY, in case the latter was bound to IscS in the same way it is within the apo ternary complex. In this case, clashes would be experienced also between CyaY and the IscS catalytic loop. These unfavourable contacts would thus make impossible for IscU or the IscS catalytic loop to assume an orientation such to allow C328 to coordinate the [Fe-S] cluster, as observed in the *A. fulgidus* structure.

Intriguingly, as we previously pointed out, several independent groups observed functional discrepancies between prokaryotic and eukaryotic frataxin, indicating that human frataxin more likely acts as an allosteric activator, rather than an inhibitor, of the [Fe-S] biosynthesis. However it was demonstrated by *in-vitro* studies that CyaY and FXN can be interchanged in the human and *E. coli* complexes without leading to an inversion of their effects on the catalytic rates of the reaction.[334]

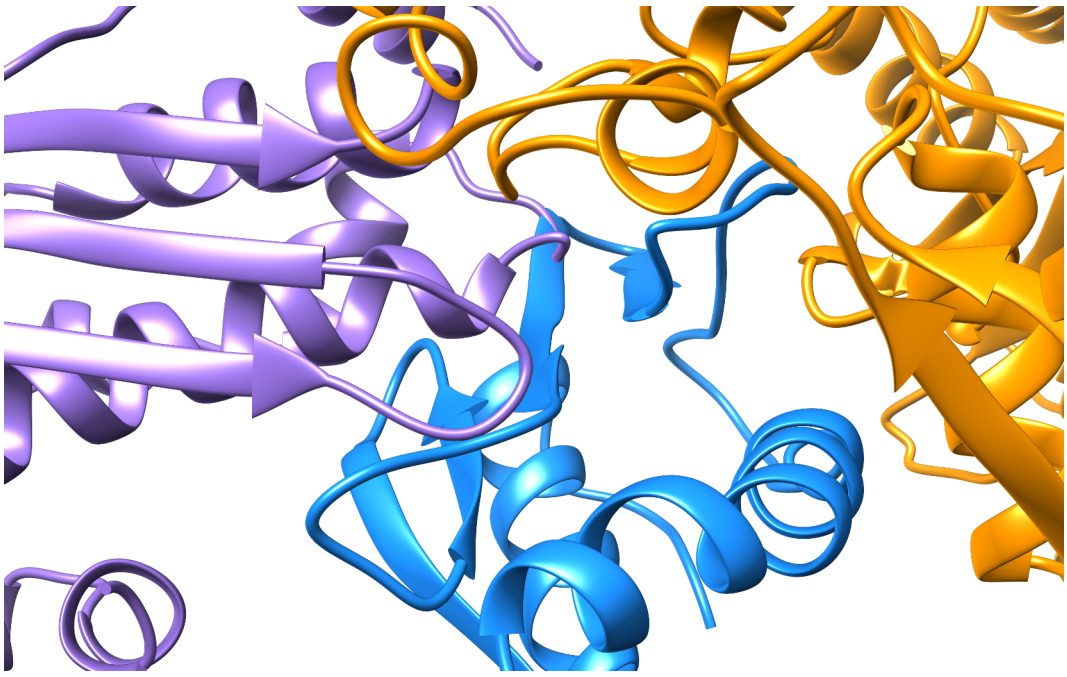


Figure 7.17: Possible steric clashes between CyaY and [Fe-S] loaded IscU in a potential holo ternary complex. The *A. fulgidus* complex structure was superimposed onto the average structure of the ternary complex, as issuing from our MD simulation, by fitting onto the IscS dimer (orange). Note how several steric clashes occur between the CyaY protomer in the ternary complex (blue) and the IscU protomer in the the *A. fulgidus* complex (purple) within this putative holo ternary complex. All proteins are depicted as cartoons.

Moreover, both CyaY and FXN can at least partially complement strains of yeast lacking the frataxin ortholog Yfh1,[355,356] again implying similar functions. Altogether, these studies indicate that the kind of frataxin allosteric modulation doesn't depend on the frataxin homolog itself, but is rather dictated by intrinsic differences, likely accumulated during evolution, in the cysteine desulphurases. Indeed, despite the relatively high sequence identity (~60%), the much lower catalytic activity of Nfs1 as compared to IscS would greatly benefit by a frataxin acting as positive regulator for both the persulphide formation and [Fe-S] cluster assembly.[324,341] Another possible difference between prokaryotic and eukaryotic desulphurases lies in the fact that the eukaryotic protein necessarily requires one more component, Isd11,[357-359] which is thought to stabilize the

Nfs1 dimer. The presence of this additional component, which is absent in prokaryotes, may compel a slightly different arrangement of frataxin in the human quaternary complex, though this could be unlikely, as it was shown that both CyaY and frataxin share similar interaction surfaces on their respective desulphurases.[351,352] However, such a rather puzzling issue will have to await until further investigations are carried out and, ultimately, the structural models of human ternary and quaternary complexes are determined to be untangled.

7.4 Conclusions

The bacterial proteins IscS and IscU are essential components of [Fe-S] protein biogenesis, an essential process for cell survival and very well conserved along the whole evolutionary tree. These proteins together form a complex where [Fe-S] clusters are assembled, both *in-vitro* and *in-vivo*, and have close orthologs in eukaryotes, up to primates, making them excellent model systems for the more complex eukaryotic machinery. In the present thorough *in-silico* study, we have unravelled a few interesting details concerning the mechanism through which this complicated protein complex works. Our results have shown that the shuttling motion through which the IscS catalytic loop was proposed to carry out its sulphur transferring role should occur via a conformational transition featuring an increase in compactness and the formation of a small 3_{10}-helix involving C328 and its flanking residues. Moreover, we have also shown that the IscS-IscU complex doesn't behave as a compact unchanging entity, but instead, the complex is held together only via loose interactions at the binding interface, as reflected by the residual flexibility of IscU monomers as a whole upon formation of the complex. Such flexibility could be of functional importance, by helping to bring the IscU active site cysteines closer to C328, thereby complementing the shuttling motion of the IscS catalytic loop during sulphur transfer.

The functioning of the IscS-IscU complex is regulated by the ancillary protein CyaY, ortholog of human frataxin, which in bacteria acts as a negative allosteric modulator of [Fe-S] cluster biosynthesis. Our results on the IscS-IscU-CyaY ternary complex have shown that the presence of CyaY strengthens the binding between IscS and IscU, stabilizing the fluctuations of the whole complex. As a result, the IscU monomers display a relatively lower mobility and the inherent flexibility of the IscS catalytic loop is almost completely abolished. Altogether, the

evidences we have provided within this work are in good agreement with former experimental data and substantiate the previously presented structural model of the IscS-IscU-CyaY complex. In addition, we have also identified several residues on the IscS and CyaY interacting surfaces, never investigated so far, that may potentially influence or abolish the binding or even the catalytic activity of the complex.

We believe that this study provides useful new insights into the mechanism through which the IscS-IscU complex works and suggests a plausible explanation for the inhibitory mechanism of CyaY. Certainly, further studies will be required to shed additional light on the catalytic mechanism of the complex or to assess whether CyaY can also impinge on other steps of the biosynthesis, such as iron uptake or persulphide reduction on C328 to yield inorganic sulphide (S^{2-}). Another important and very fascinating task would be digging into the differences between human and *E. coli* complexes and how these differences control, in an opposite fashion, [Fe-S] cluster assembly in prokaryotes and eukaryotes. The information provided in our work will hopefully inspire new studies aimed at solving these still open questions, in order to better understand both one of the most essential metabolic pathways and the molecular basis of severe diseases, such as Friederich's ataxia.

Chapter 8

Conclusions and perspectives

Molecular modelling based approaches are nowadays widely used to gain insights into the molecular basis of protein biophysical properties and cellular functions, within a biological-like environment. In the present thesis, molecular dynamics simulation and free energy calculation techniques have been applied to the study of the dynamics and of the working mechanism of naturally occurring biological systems and engineered protein based molecular nanodevices.

The simulations of the 5-HT_3A receptor are generally in good agreement with recent and previous experimental findings and computational studies performed on homologue receptors. In addition, further insights into the ion translocation mechanism through this receptor can also be drawn. The discovery of the lateral intersubunit channels is particularly interesting in this respect, since they are located within a region of the protein which is thought to be pivotal for conveying the energy arising from the ligand binding event to the transmembrane gate, eventually leading to the opening of the pore. Moreover, the fact that these lateral channels seem to be shared also by other pLGICs, and particularly only by eukaryotic members, makes these structural features even more intriguing and could stimulate further theoretical and experimental investigation on this and other related receptors. The identification of few residues on the main channel internal walls as potentially influencing the ion migration process will be useful for designing more focused experiments, so as to attain a deeper comprehension about how the 5-HT_3AR functions. Finally, the investigation into the role of the arginine triplet within the ICD has nicely reconciled experimental conductance

observations with a physically sound molecular explanation based on electrostatics and sterical encumbrance, illuminating the mechanism through which they act as determinants of the 5-HT_3A low channel conductance.

The simulations of the engineered MscL have shed light on the first stages of the light activated gating mechanism artificially introduced within this gated ion channel. The evidence that the channel gating relies upon a charge unmasking mechanism and consequent hydrophobicity breaking in the middle of the pore allows to envisage that it would be possible to obtain a much broader range of light gated nanodevices. Indeed, a hydrophobic gating mechanism seems to be shared also by many other gated ion channels, and not only by MscL. Moreover, this evidence suggests the hypothesis that replacements with other polar amino acids may also be suitable to open the pore, thus further enriching the arsenal of available functionalizations. A broader toolbox of light triggered nanodevices may exploit the benefits of different kinds of photoactive functionalizations in two ways. On the one hand the possibility to uncover different chemical groups, while still inducing the opening of the channel, could allow to employ a greater number of photoactive moieties, thus allowing to synthetize devices sensitive to a wider spectrum of wavelengths. On the other hand, the possibility to finely tune the polarity of the chemical groups could allow to achieve a better control on the final desired size of the pore in the conductive state. This, along with the chance of a subunit specific triggering, could possibly even lead to a finer control on the transport properties, such as the release kinetics, and prove useful in several instances, e.g. in case of drug delivery oriented nanodevices.

Simulations of the engineered α-Hemolysin have shown the viability of the application of molecular modelling approaches to the *de-novo* rational design of bio-inspired nanopores with controllable properties. Thanks to the tethering of an appropriately conceived azobenzene based photo-switch at a carefully determined location on the rigid α-Hemolysin protein scaffold, a light controllable nanopore has been devised, whose promising capabilities have been evidenced in this study. In particular, the possibility of blocking the passage of medium size molecules, such as a small DNA fragment, makes it as a prototype system suitable for many diverse applications, such as in analytical chemistry or drug delivery. A particularly appealing application is also in the increasingly popular field of nanopore based DNA sequencing, where α-Hemolysin has already demonstrated its high potential. The future developments are many. For instance, the modification of the “on/off” molecular photo-switch presented here in terms of its

chemical nature, length of the linker, volume and net charge of the head group may allow to achieve desired selective transport properties, suitable for a particular application. In this respect, a thorough and systematic computational assessment of the chemical features of the photo-switchable moiety could be carried out prior to undertaking a long and expansive synthetic chemistry campaign, aiding in selecting only a handful of promising candidates. The experimental validation of the theoretical predictions will then help to obtain a more clear picture of the structure-property relationships, creating a sort of "integrative approach" between experimental and computational data. This may prove of great benefit for reducing the optimization costs for the process of identification of a successful photo-sensitive group. Finally, other unspecific and non-gated biological pores may be screened, possibly also exploring different functionalization sites on the scaffold proteins, thereby widely expanding the set of molecular devices that can be conceived.

The simulation of the IscS-IscU system, both in presence and absence of the allosteric inhibitor CyaY, has provided new important insights into the working mechanism of such a vital protein complex, adding one more piece to the [Fe-S] protein biogenesis intricate puzzle. The characteristic IscU movement with respect to IscS, likely brought about by a relatively "soft" binding interface, is particularly interesting in this respect. Indeed such movement possibly elucidates how IscU switches between two different orientations within the complex, based on the presence of the [Fe-S] cluster. The portrayal of IscU as a mere static platform appointed to receive the components needed to build up [Fe-S] clusters should then be revised in favour of a closer cooperation between such scaffold-protein and the desulphurase. An active involvement of IscU may also help to provide a plausible explanation for the CyaY inhibitory mechanism in bacterial systems, which would be compatible with that based on abolishing the IscS catalytic loop shuttling motion observed in the CyaY free system. Such explanation would help to rationalize both why the presence of CyaY is apparently incompatible with the orientation assumed by the IscU holo-form within the complex and why CyaY seems to affect the IscU movement mentioned above. Unfortunately, the yet poorly characterized difference between prokaryotic and eukaryotic protein complexes does not allow to directly exploit the findings of this work to develop a therapy for the many severe human diseases related to defects in [Fe-S] protein biogenesis systems. Nonetheless, some of them may still prove useful to pursue a different, but not less worthy, goal. Indeed, as the

presence of the ISC system is usually very well conserved, the availability of effective pharmacological tools targeting the IscS-CyaY interaction could be extremely useful to achieve a deeper understanding of the cellular role of the interaction between CyaY and the IscS-IscU complex (or their orthologues). By providing hints about which residues, if mutated, are more likely able to influence the binding, the characterization of the interface between IscS and CyaY presented here could foster pilot experimental studies specifically focused at identifying differences between systems belonging to different organisms. Once these differences will have been determined, it will be possible to start a drug discovery campaign aimed at achieving successful PPIs inhibitors that can serve as potentially selective biochemical probes .

In conclusion, the work presented here overall highlights that MD simulations are twice as valuable tools to the study of biomolecular system. On the one hand, they can be used *ex-ante* to make predictions and provide useful suggestions aimed at helping the design of new experiments. On the other hand, they can be used *ex-post* to provide detailed insights, otherwise unavailable, into the working mechanism of biologically and biotechnologically relevant systems, thereby complementing and aiding to rationalize experimental observations. The diversity of the systems investigated within my PhD thesis work emphasizes the power and versatility of computational chemistry in modern chemical and biological research.

Acknowledgements

First of all, I thank my tutor, Dr. Giuseppe Brancato, for his guidance and for the invaluable assistance and support he provided me with during my PhD.

I am particularly grateful to Prof. Vincenzo Barone for allowing me to join his computational laboratory in "*Scuola Normale Superiore*" and for giving me the opportunity to join other groups during my training as a PhD to broaden my scientific knowledge.

I am also very grateful to Prof. Ettore Novellino for giving me the opportunity to join his group in Naples during my PhD.

A special thank goes to Balasubramanian Chandramouli from *Scuola Normale Superiore* for his constant technical, scientific and human support. I also thank Prof. Luciana Marinelli and Dr Francesco Saverio di Leva from *University of Naples "Federico II",* Dr Sandro Cosconati from "*Second University of Naples*" and to Dr Vittorio Limongelli from "*U.S.I.*, Lugano" for their precious advices and suggestions.

Furthermore, I want to mention Prof. Annalisa Pastore from "*King's College*, London" for her precious contribution to the work presented in Chapter 7.

Finally, I want to thank all my lab buddies, and especially Matteo Piccardo, Danilo Calderini, Teresa Fornaro and Giordano Mancini for their sincere friendship.

Abbreviations and Symbols

Å	Ångstrom
ΔA	difference in free energy
β	thermodynamic beta
C_α	alpha carbon atom
Da	Dalton
ΔG	difference in Gibbs free energy
η	weighting factor for biasing phase space exploration
f	microscopic phase space function
$\mathcal{F}$	Faraday constant
$\overline{F}_\xi$	average force along a generic collective variable
F_μ^k	μ-th instantaneous force sample in bin k in ABF simulations
H	Hamiltonian of a system
k_B	Boltzmann constant
K_d	dissociation constant
ξ	generic collective variable
$\boldsymbol{\Xi}$	transformation from Cartesian to generalized coordinates
J	Jacobian matrix
L	generic thermodynamical observable

N_0	threshold number of samples beyond which applying the average force in ABF simulations
N_t^k	total number of samples accrued in bin k at time t during ABF simulations
$\boldsymbol{x}$	vector containing atomic positions
$\boldsymbol{p}$	vector containing atomic momenta
P	probability distribution function
P_i	permeability of ion i
R	gas constant
q	partial atomic charge
q_i	generic generalized coordinate
Q	ionic charge
U	potential energy of a system
V	generic biasing potential
V_m	membrane potential
Z	molecular partition function
Ω	domain of the accessible micro states for the energy levels compatible with a given macro state
2D	2-dimensional
3D	3-dimensional
5-HT3AR	5-Hydroxytriptamine Receptor of type 3A
ABF	Adaptive Biasing Force
αHL	α-Hemolysin from *Staphylococcus aureus*
ATP	Adenosine Triphosphate
Cryo-EM	Cryo-Electron Microscopy
dA	deoxyadenosine
DCC	Dynamic Cross Correlation

DNA	Deoxyribonucleic acid
DOF	Degree of Freedom
ED	Essential Dynamics
ESP	Electrostatic Potential
FXN	Frataxin
GHK	Goldman-Hodgkin-Katz
GIC	Gated Ion Channels
ISC	Iron Sulphur Cluster
LJ	Lennard-Jones
MD	Molecular Dynamics
MM	Molecular Mechanics
mRNA	messenger Ribonucleic Acid
MscL	Mechanosensitive Channel of Large conductance
MTS	Methanethiosulfonate
NMR	Nuclear Magnetic Resonance
NPT	Isobaric-Isothermal ensemble, characterized by a fixed number of particles (N), pressure (P) and Temperature (T)
NVT	Canonical ensemble, characterized by a fixed number of particles (N), volume (V) and Temperature (T)
PBC	Periodic Boundary Conditions
PCA	Principal Component Analysis
PDB	Protein Data Bank
PES	Potential Energy Surface
PFAP	Pore Forming Antimicrobial Peptide
PFT	Pore Forming Toxin
pLGIC	pentameric Ligand-Gated Ion Channels
PLP	Pyridoxal-phosphate

PME	Particle Mesh Ewald
PMF	Potential of Mean Force
PPI	Protein-Protein Interaction
QM	Quantum Mechanics
RESP	Restricted Electrostatic Potential
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
ROS	Reactive Oxygen Species
SC_4	Sulphonato-Calix[4]arene
SMD	Steered Molecular Dynamics
TI	Thermodynamic Integration
TM	Transmembrane
tRNA	transfer Ribonucleic Acid
vdW	van der Waals
WT	Wild Type

Bibliography

1. Cannon WB (1926) Physiological Regulation of Normal States: Some Tentative Postulates Concerning Biological Homeostatics. Paris: Editions Medicales.

2. Marieb EN (2009) Essentials of human anatomy & physiology. San Francisco, CA: Pearson/Benjamin Cummings. 632 p.

3. Kotas ME, Medzhitov R (2015) Homeostasis, inflammation, and disease susceptibility. Cell 160: 816-827.

4. Waldron KJ, Rutherford JC, Ford D, Robinson NJ (2009) Metalloproteins and metal sensing. Nature 460: 823-830.

5. Andreini C, Bertini I, Cavallaro G, Najmanovich RJ, Thornton JM (2009) Structural analysis of metal sites in proteins: non-heme iron sites as a case study. Journal of Molecular Biology 388: 356-380.

6. Guengerich FP (2009) Thematic Series: Metals in Biology. Journal of Biological Chemistry 284: 709-709.

7. Finney LA, O'Halloran TV (2003) Transition metal speciation in the cell: insights from the chemistry of metal ion receptors. Science (New York, NY) 300: 931-936.

8. Andreini C, Banci L, Bertini I, Rosato A (2006) Counting the Zinc-Proteins Encoded in the Human Genome. Journal of Proteome Research 5: 196-201.

9. Dlouhy AC, Outten CE (2013) The Iron Metallome in Eukaryotic Organisms. In: Banci L, editor. Metallomics and the Cell. Dordrecht: Springer Netherlands. pp. 241-278.

10. Hulpiau P, van Roy F (2009) Molecular evolution of the cadherin superfamily. The International Journal of Biochemistry & Cell Biology 41: 349-369.

11. Iwamoto DV, Calderwood DA (2015) Regulation of integrin-mediated adhesions. Current Opinion in Cell Biology 36: 41-47.

12. Brini M, Calì T, Ottolini D, Carafoli E (2013) Intracellular Calcium Homeostasis and Signaling. In: Banci L, editor. Metallomics and the Cell. Dordrecht: Springer Netherlands. pp. 119-168.

13. Sychrová H (2004) Yeast as a model organism to study transport and homeostasis of alkali metal cations. Physiological Research / Academia Scientiarum Bohemoslovaca 53 Suppl 1: S91-98.

14. Steinbach HB, Spiegelman S (1943) The sodium and potassium balance in squid nerve axoplasm. Journal of Cellular and Comparative Physiology 22: 187-196.

15. (2002) General Principles of Cell Communication. In: Alberts B, editor. Molecular biology of the cell. 4th ed ed. New York: Garland Science.

16. Hodgkin AL (1951) THE IONIC BASIS OF ELECTRICAL ACTIVITY IN NERVE AND MUSCLE. Biological Reviews 26: 339-409.

17. Goldman DE (1943) POTENTIAL, IMPEDANCE, AND RECTIFICATION IN MEMBRANES. The Journal of General Physiology 27: 37-60.

18. Hille B (2001) Ion channels of excitable membranes. Sunderland, Mass: Sinauer. 814 p.

19. Gouaux E (2005) Principles of Selective Ion Transport in Channels and Pumps. Science 310: 1461-1465.

20. Tombola F, Pathak MM, Isacoff EY (2006) How Does Voltage Open an Ion Channel? Annual Review of Cell and Developmental Biology 22: 23-52.

21. Ashcroft FM (2006) From molecule to malady. Nature 440: 440-447.

22. Zaydman MA, Silva JR, Cui J (2012) Ion Channel Associated Diseases: Overview of Molecular Mechanisms. Chemical Reviews 112: 6319-6333.

23. Menestrina G, Dalla Serra M (2003) Pore-forming Peptides and Protein Toxins. 29 West 35th Street, New York, NY 10001: CRC Press. 360 p.

24. Morgan PJ, Hyman SC, Rowe AJ, Mitchell TJ, Andrew PW, Saibil HR (1995) Subunit organisation and symmetry of pore-forming, oligomeric pneumolysin. FEBS letters 371: 77-80.

25. Behrends JC (2012) Evolution of the Ion Channel Concept: The Historical Perspective. Chemical Reviews 112: 6218-6226.

26. Haswell ES, Phillips R, Rees DC (2011) Mechanosensitive channels: what can they do and how do they do it? Structure (London, England: 1993) 19: 1356-1369.

27. Dhaka A, Viswanath V, Patapoutian A (2006) TRP ION CHANNELS AND TEMPERATURE SENSATION. Annual Review of Neuroscience 29: 135-161.

28. Holzer P (2009) Acid-Sensitive Ion Channels and Receptors. In: Canning BJ, Spina D, editors. Sensory Nerves. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 283-332.

29. Nagel G, Ollig D, Fuhrmann M, Kateriya S, Musti AM, Bamberg E, et al. (2002) Channelrhodopsin-1: a light-gated proton channel in green algae. Science (New York, NY) 296: 2395-2398.

30. Lemoine D, Jiang R, Taly A, Chataigneau T, Specht A, Grutter T (2012) Ligand-Gated Ion Channels: New Insights into Neurological Disorders and Ligand Recognition. Chemical Reviews 112: 6285-6318.

31. Catterall WA (2010) Ion Channel Voltage Sensors: Structure, Function, and Pathophysiology. Neuron 67: 915-928.

32. Perozo E, Rees DC (2003) Structure and mechanism in prokaryotic mechanosensitive channels. Current Opinion in Structural Biology 13: 432-442.

33. Thompson AJ, Lester HA, Lummis SCR (2010) The structural basis of function in Cys-loop receptors. Quarterly Reviews of Biophysics 43: 449-499.

34. Lester HA, Dibas MI, Dahan DS, Leite JF, Dougherty DA (2004) Cys-loop receptors: new twists and turns. Trends in Neurosciences 27: 329-336.

35. Beckstein O, Sansom MSP (2006) A hydrophobic gate in an ion channel: the closed state of the nicotinic acetylcholine receptor. Physical Biology 3: 147-159.

36. Beckstein O, Sansom MSP (2003) Liquid-vapor oscillations of water in hydrophobic nanopores. Proceedings of the National Academy of Sciences 100: 7063-7068.

37. Rasaiah JC, Garde S, Hummer G (2008) Water in Nonpolar Confinement: From Nanotubes to Proteins and Beyond [*]. Annual Review of Physical Chemistry 59: 713-740.

38. Aryal P, Sansom MSP, Tucker SJ (2015) Hydrophobic Gating in Ion Channels. Journal of Molecular Biology 427: 121-130.

39. Jensen MO, Jogini V, Borhani DW, Leffler AE, Dror RO, Shaw DE (2012) Mechanism of Voltage Gating in Potassium Channels. Science 336: 229-233.

40. Anishkin A, Akitake B, Kamaraju K, Chiang CS, Sukharev S (2010) Hydration properties of mechanosensitive channel pores define the energetics of gating. Journal of Physics Condensed Matter: An Institute of Physics Journal 22: 454120.

41. Sauguet L, Shahsavar A, Delarue M (2015) Crystallographic studies of pharmacological sites in pentameric ligand-gated ion channels. Biochimica et Biophysica Acta (BBA) - General Subjects 1850: 511-523.

42. Chang G, Spencer RH, Lee AT, Barclay MT, Rees DC (1998) Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel. Science (New York, NY) 282: 2220-2226.

43. Dutzler R, Campbell EB, Cadene M, Chait BT, MacKinnon R (2002) X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. Nature 415: 287-294.

44. Bass RB, Strop P, Barclay M, Rees DC (2002) Crystal structure of Escherichia coli MscS, a voltage-modulated and mechanosensitive channel. Science (New York, NY) 298: 1582-1587.

45. Uysal S, Vásquez V, Tereshko V, Esaki K, Fellouse FA, Sidhu SS, et al. (2009) Crystal structure of full-length KcsA in its closed conformation. Proceedings of the National Academy of Sciences of the United States of America 106: 6644-6649.

46. Long SB, Tao X, Campbell EB, MacKinnon R (2007) Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment. Nature 450: 376-382.

47. Kawate T, Michel JC, Birdsong WT, Gouaux E (2009) Crystal structure of the ATP-gated P2X4 ion channel in the closed state. Nature 460: 592-598.

48. Hassaine G, Deluz C, Grasso L, Wyss R, Tol MB, Hovius R, et al. (2014) X-ray structure of the mouse serotonin 5-HT3 receptor. Nature 512: 276-281.

49. Corringer P-J, Poitevin F, Prevost Marie S, Sauguet L, Delarue M, Changeux J-P (2012) Structure and Pharmacology of Pentameric Receptor Channels: From Bacteria to Brain. Structure 20: 941-956.

50. Iscla I, Blount P (2012) Sensing and Responding to Membrane Tension: The Bacterial MscL Channel as a Model System. Biophysical Journal 103: 169-174.

51. Bernheimer AW (1970) Cytolytic toxins of bacteria. Microbial toxins. Academic ed. New York: Ajl S, Kadis S, Montie TC. pp. 183-212.

52. Thelestam M, Möllby R (1979) Classification of microbial, plant and animal cytolysins based on their membrane-damaging effects on human fibroblasts. Biochimica et Biophysica Acta (BBA) - Biomembranes 557: 156-169.

53. Mims CA, Nash A, Stephen J (2001) Mims' Pathogenesis of infectious disease. San Diego: Academic Press, c2001. 474 p.

54. Sato N, Kurotaki H, Watanabe T, Mikami T, Matsumoto T (1998) Use of hemoglobin as an iron source by Bacillus cereus. Biological & Pharmaceutical Bulletin 21: 311-314.

55. van der Goot FG (2000) Plasticity of the transmembrane beta-barrel. Trends in Microbiology 8: 89-90.

56. Gilbert RJ, Jiménez JL, Chen S, Tickle IJ, Rossjohn J, Parker M, et al. (1999) Two structural transitions in membrane pore formation by pneumolysin, the pore-forming toxin of Streptococcus pneumoniae. Cell 97: 647-655.

57. Bernheimer AW, Rudy B (1986) Interactions between membranes and cytolytic peptides. Biochimica Et Biophysica Acta 864: 123-141.

58. Townsley LE, Tucker WA, Sham S, Hinton JF (2001) Structures of gramicidins A, B, and C incorporated into sodium dodecyl sulfate micelles. Biochemistry 40: 11676-11686.

59. Song C, Weichbrodt C, Salnikov ES, Dynowski M, Forsberg BO, Bechinger B, et al. (2013) Crystal structure and functional mechanism of a human antimicrobial membrane channel. Proceedings of the National Academy of Sciences 110: 4586-4591.

60. Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE (1996) Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. Science (New York, NY) 274: 1859-1866.

61. Mueller M, Grauschopf U, Maier T, Glockshuber R, Ban N (2009) The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. Nature 459: 726-730.

62. Tilley SJ, Orlova EV, Gilbert RJC, Andrew PW, Saibil HR (2005) Structural Basis of Pore Formation by the Bacterial Toxin Pneumolysin. Cell 121: 247-256.

63. Mignogna G, Simmaco M, Barra D (1998) Occurrence and function of D-amino acid-containing peptides and proteins: antimicrobial peptides. EXS 85: 29-36.

64. Hwang PM, Vogel HJ (1998) Structure-function relationships of antimicrobial peptides. Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire 76: 235-246.

65. Oren Z, Shai Y (1998) Mode of action of linear amphipathic alpha-helical antimicrobial peptides. Biopolymers 47: 451-463.

66. Matsuzaki K (1998) Magainins as paradigm for the mode of action of pore forming polypeptides. Biochimica Et Biophysica Acta 1376: 391-400.

67. Szymański W, Beierle JM, Kistemaker HAV, Velema WA, Feringa BL (2013) Reversible Photocontrol of Biological Systems by the Incorporation of Molecular Photoswitches. Chemical Reviews 113: 6114-6178.

68. Majd S, Yusko EC, Billeh YN, Macrae MX, Yang J, Mayer M (2010) Applications of biological pores in nanomedicine, sensing, and nanoelectronics. Current Opinion in Biotechnology 21: 439-476.

69. Kasianowicz JJ, Brandin E, Branton D, Deamer DW (1996) Characterization of individual polynucleotide molecules using a membrane channel. Proceedings of the National Academy of Sciences of the United States of America 93: 13770-13773.

70. Grosse W, Essen L-O, Koert U (2011) Strategies and perspectives in ion-channel engineering. Chembiochem: A European Journal of Chemical Biology 12: 830-839.

71. Reiner JE, Balijepalli A, Robertson JWF, Campbell J, Suehle J, Kasianowicz JJ (2012) Disease detection and management via single nanopore-based sensors. Chemical Reviews 112: 6431-6451.

72. Banghart MR, Volgraf M, Trauner D (2006) Engineering light-gated ion channels. Biochemistry 45: 15129-15141.

73. Volgraf M, Gorostiza P, Numano R, Kramer RH, Isacoff EY, Trauner D (2006) Allosteric control of an ionotropic glutamate receptor with an optical switch. Nature Chemical Biology 2: 47-52.

74. Stankovic CJ, Heinemann SH, Schreiber SL (1991) Photo-modulated ion channels based on covalently linked gramicidins. Biochimica Et Biophysica Acta 1061: 163-170.

75. Loudwig S, Bayley H (2006) Photoisomerization of an individual azobenzene molecule in water: an on-off switch triggered by light at a fixed wavelength. Journal of the American Chemical Society 128: 12404-12405.

76. Lester HA, Chang HW (1977) Response of acetylcholine receptors to rapid photochemically produced increases in agonist concentration. Nature 266: 373-374.

77. Koçer A, Walko M, Meijberg W, Feringa BL (2005) A light-actuated nanovalve derived from a channel protein. Science (New York, NY) 309: 755-758.

78. Howorka S, Siwy Z (2009) Nanopore analytics: sensing of single molecules. Chemical Society Reviews 38: 2360-2384.

79. Frey PA, Reed GH (2012) The ubiquity of iron. ACS chemical biology 7: 1477-1481.

80. Imlay JA (2013) The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. Nature Reviews Microbiology 11: 443-454.

81. Kell DB (2009) Iron behaving badly: inappropriate iron chelation as a major contributor to the aetiology of vascular and other progressive inflammatory and degenerative diseases. BMC medical genomics 2: 2.

82. Dixon SJ, Stockwell BR (2013) The role of iron and reactive oxygen species in cell death. Nature Chemical Biology 10: 9-17.

83. Kosman DJ (2010) Redox Cycling in Iron Uptake, Efflux, and Trafficking. Journal of Biological Chemistry 285: 26729-26735.

84. Crichton RR, Charloteaux-Wauters M (1987) Iron transport and storage. European Journal of Biochemistry 164: 485-506.

85. Cheng Y, Zak O, Aisen P, Harrison SC, Walz T (2004) Structure of the Human Transferrin Receptor-Transferrin Complex. Cell 116: 565-576.

86. Andrews SC, Harrison PM, Yewdall SJ, Arosio P, Levi S, Bottke W, et al. (1992) Structure, function, and evolution of ferritins. Journal of Inorganic Biochemistry 47: 161-174.

87. Beinert H (1997) Iron-Sulfur Clusters: Nature's Modular, Multipurpose Structures. Science 277: 653-659.

88. Johnson DC, Dean DR, Smith AD, Johnson MK (2005) STRUCTURE, FUNCTION, AND FORMATION OF BIOLOGICAL IRON-SULFUR CLUSTERS. Annual Review of Biochemistry 74: 247-281.

89. Tagawa K, Arnon DI (1962) Ferredoxins as electron carriers in photosynthesis and in the biological production and consumption of hydrogen gas. Nature 195: 537-543.

90. Kuo CF, McRee DE, Fisher CL, O'Handley SF, Cunningham RP, Tainer JA (1992) Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. Science (New York, NY) 258: 434-440.

91. Guan Y, Manuel RC, Arvai AS, Parikh SS, Mol CD, Miller JH, et al. (1998) MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. Nature Structural Biology 5: 1058-1064.

92. Kiley PJ, Beinert H (2003) The role of Fe-S proteins in sensing and regulation in bacteria. Current Opinion in Microbiology 6: 181-185.

93. Demple B, Ding H, Jorgensen M (2002) Escherichia coli SoxR protein: sensor/transducer of oxidative stress and nitric oxide. Methods in Enzymology 348: 355-364.

94. Beinert H, Kennedy MC, Stout CD (1996) Aconitase as Ironminus signSulfur Protein, Enzyme, and Iron-Regulatory Protein. Chemical Reviews 96: 2335-2374.

95. Ye H, Rouault TA (2010) Human Iron−Sulfur Cluster Assembly, Cellular Iron Homeostasis, and Disease. Biochemistry 49: 4945-4956.

96. Barras F, Loiseau L, Py B (2005) How Escherichia coli and Saccharomyces cerevisiae Build Fe/S Proteins. Advances in Microbial Physiology: Elsevier. pp. 41-101.

97. Muhlenhoff U (2003) Components involved in assembly and dislocation of iron-sulfur clusters on the scaffold protein Isu1p. The EMBO Journal 22: 4815-4825.

98. Xu XM, Møller SG (2008) Iron-Sulfur Cluster Biogenesis Systems and their Crosstalk. ChemBioChem 9: 2355-2362.

99. Jacobson MR, Cash VL, Weiss MC, Laird NF, Newton WE, Dean DR (1989) Biochemical and genetic analysis of the nifUSVWZM cluster from Azotobacter vinelandii. Molecular & general genetics: MGG 219: 49-57.

100. Zheng M, Wang X, Templeton LJ, Smulski DR, LaRossa RA, Storz G (2001) DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide. Journal of Bacteriology 183: 4562-4570.

101. Blanc B, Gerez C, Ollagnier de Choudens S (2015) Assembly of Fe/S proteins in bacterial systems. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1853: 1436-1447.

102. Chandler D (1987) Introduction to modern statistical mechanics. New York: Oxford University Press. 274 p.

103. Tuckerman ME (2010) Statistical mechanics: theory and molecular simulation. Oxford ; New York: Oxford University Press. 696 p.

104. Allen MP, Tildesley DJ (1987) Computer simulation of liquids. Oxford [England] : New York: Clarendon Press ; Oxford University Press. 385 p.

105. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics 65: 712-725.

106. MacKerell AD, Banavali N, Foloppe N (2000) Development and current status of the CHARMM force field for nucleic acids. Biopolymers 56: 257-265.

107. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics 98: 10089.

108. (2007) Free energy calculations: theory and applications in chemistry and biology; Chipot C, Pohorille A, editors. Berlin ; New York: Springer. 517 p.

109. (2008) Rugged free energy landscapes: common computational approaches in spin glasses, structural glasses, and biological macromolecules; Janke W, editor. Berlin ; New York: Springer. 412 p.

110. Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, et al. (2011) Structure and function of an irreversible agonist-β2 adrenoceptor complex. Nature 469: 236-240.

111. Dror RO, Green HF, Valant C, Borhani DW, Valcourt JR, Pan AC, et al. (2013) Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. Nature.

112. Kohlhoff KJ, Shukla D, Lawrenz M, Bowman GR, Konerding DE, Belov D, et al. (2013) Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. Nature Chemistry 6: 15-21.

113. Bernardi RC, Melo MCR, Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochimica et Biophysica Acta (BBA) - General Subjects 1850: 872-877.

114. Valleau JP (1972) Monte Carlo Estimation of the Free Energy by Multistage Sampling. The Journal of Chemical Physics 57: 5457.

115. Srinivasan R (2002) Importance Sampling. Berlin, Heidelberg: Springer Berlin Heidelberg.

116. Chipot C, Pohorille A (1998) Conformational equilibria of terminally blocked single amino acids at the water-hexane interface. A molecular dynamics study. The Journal of Physical Chemistry B 102: 281-290.

117. Born M (1920) Volumen und Hydratationswärme der Ionen. Zeitschrift für Physik 1: 45-48.

118. Kirkwood JG (1935) Statistical Mechanics of Fluid Mixtures. The Journal of Chemical Physics 3: 300.

119. Carter EA, Ciccotti G, Hynes JT, Kapral R (1989) Constrained reaction coordinate dynamics for the simulation of rare events. Chemical Physics Letters 156: 472-477.

120. Straatsma TP, McCammon JA (1991) Multiconfiguration thermodynamic integration. The Journal of Chemical Physics 95: 1175.

121. Sprik M, Ciccotti G (1998) Free energy from constrained molecular dynamics. The Journal of Chemical Physics 109: 7737.

122. Hénin Jr, Fiorin G, Chipot C, Klein ML (2010) Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables. Journal of Chemical Theory and Computation 6: 35-47.

123. den Otter WK (2000) Thermodynamic integration of the free energy along a reaction coordinate in Cartesian coordinates. The Journal of Chemical Physics 112: 7283.

124. Ciccotti G, Kapral R, Vanden-Eijnden E (2005) Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics. Chemphyschem: A European Journal of Chemical Physics and Physical Chemistry 6: 1809-1814.

125. Lelièvre T, Stoltz G, Rousset M (2010) Free energy computations: a mathematical perspective. London ; Hackensack, N.J: Imperial College Press. 458 p.

126. Darve E, Pohorille A (2001) Calculating free energies using average force. The Journal of Chemical Physics 115: 9169.

127. Hénin Jrm, Chipot C (2004) Overcoming free energy barriers using unconstrained molecular dynamics simulations. The Journal of Chemical Physics 121: 2904.

128. Darve E, Rodríguez-Gómez D, Pohorille A (2008) Adaptive biasing force method for scalar and vector free energy calculations. The Journal of Chemical Physics 128: 144120.

129. Comer J, Gumbart JC, Hénin J, Lelièvre T, Pohorille A, Chipot C (2015) The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. The Journal of Physical Chemistry B 119: 1129-1151.

130. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577-2637.

131. Roe DR, Cheatham TE (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation 9: 3084-3095.

132. McCammon JA, Harvey SC (1987) Dynamics of proteins and nucleic acids. Cambridge [Cambridgeshire] ; New York: Cambridge University Press. 234 p.

133. Hayward S, Groot BL (2008) Normal Modes and Essential Dynamics. In: Kukol A, editor. Molecular Modeling of Proteins. Totowa, NJ: Humana Press. pp. 89-106.

134. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proceedings of the National Academy of Sciences of the United States of America 98: 10037-10041.

135. Smart OS, Neduvelil JG, Wang X, Wallace BA, Sansom MSP (1996) HOLE: A program for the analysis of the pore dimensions of ion channel structural models. Journal of Molecular Graphics 14: 354-360.

136. Yaffe E, Fishelovitch D, Wolfson HJ, Halperin D, Nussinov R (2008) MolAxis: a server for identification of channels in macromolecules. Nucleic Acids Research 36: W210-W215.

137. Nys M, Kesters D, Ulens C (2013) Structural insights into Cys-loop receptor function and ligand recognition. Biochemical Pharmacology 86: 1042-1053.

138. Dougherty DA (2008) Cys-Loop Neuroreceptors: Structure to the Rescue? Chemical Reviews 108: 1642-1653.

139. Bouzat C, Gumilar F, Spitzmaul G, Wang H-L, Rayes D, Hansen SB, et al. (2004) Coupling of agonist binding to channel gating in an ACh-binding protein linked to an ion channel. Nature 430: 896-900.

140. Imoto K, Busch C, Sakmann B, Mishina M, Konno T, Nakai J, et al. (1988) Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance. Nature 335: 645-648.

141. Lee WY, Sine SM (2005) Principal pathway coupling agonist binding to channel gating in nicotinic receptors. Nature 438: 243-247.

142. Lummis SCR, Beene DL, Lee LW, Lester HA, Broadhurst RW, Dougherty DA (2005) Cis–trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. Nature 438: 248-252.

143. Beckstein O, Sansom MS (2006) A hydrophobic gate in an ion channel: the closed state of the nicotinic acetylcholine receptor. Phys Biol 3: 147-159.

144. Calimet N, Simoes M, Changeux JP, Karplus M, Taly A, Cecchini M (2013) PNAS Plus: From the Cover: A gating mechanism of pentameric ligand-gated ion channels. Proceedings of the National Academy of Sciences 110: E3987-E3996.

145. Cheng X, Lu B, Grant B, Law RJ, McCammon JA (2006) Channel Opening Motion of α7 Nicotinic Acetylcholine Receptor as Suggested by Normal Mode Analysis. Journal of Molecular Biology 355: 310-324.

146. Ivanov I, Cheng X, Sine SM, McCammon JA (2007) Barriers to Ion Translocation in Cationic and Anionic Receptors from the Cys-Loop Family. Journal of the American Chemical Society 129: 8217-8224.

147. Law RJ, Henchman RH, McCammon JA (2005) A gating mechanism proposed from a simulation of a human α7 nicotinic acetylcholine receptor. Proceedings of the National Academy of Sciences of the United States of America 102: 6813-6818.

148. Nury H, Poitevin F, Van Renterghem C, Changeux JP, Corringer PJ, Delarue M, et al. (2010) One-microsecond molecular dynamics simulation of channel gating in a nicotinic receptor homologue. Proceedings of the National Academy of Sciences 107: 6275-6280.

149. Xu Y, Barrantes FJ, Luo X, Chen K, Shen J, Jiang H (2005) Conformational Dynamics of the Nicotinic Acetylcholine Receptor Channel: A 35-ns Molecular Dynamics Simulation Study. Journal of the American Chemical Society 127: 1291-1299.

150. van Dijk WJ, Klaassen RV, Schuurmans M, van der Oost J, Smit AB, Sixma TK, et al. (2001) Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. Nature 411: 269-276.

151. daCosta Corrie JB, Baenziger John E (2013) Gating of Pentameric Ligand-Gated Ion Channels: Structural Insights and Ambiguities. Structure 21: 1271-1283.

152. Lee WY, Free CR, Sine SM (2009) Binding to Gating Transduction in Nicotinic Receptors: Cys-Loop Energetically Couples to Pre-M1 and M2-M3 Regions. Journal of Neuroscience 29: 3189-3199.

153. Grosman C, Zhou M, Auerbach A (2000) Mapping the conformational wave of acetylcholine receptor channel gating. Nature 403: 773-776.

154. Cymes GD, Grosman C, Auerbach A (2002) Structure of the Transition State of Gating in the Acetylcholine Receptor Channel Pore: A Φ-Value Analysis †. Biochemistry 41: 5548-5555.

155. Purohit P, Gupta S, Jadey S, Auerbach A (2013) Functional anatomy of an allosteric protein. Nature Communications 4.

156. Limapichat W, Lester HA, Dougherty DA (2010) Chemical Scale Studies of the Phe-Pro Conserved Motif in the Cys Loop of Cys Loop Receptors. Journal of Biological Chemistry 285: 8976-8984.

157. Melis C, Bussi G, Lummis SCR, Molteni C (2009) Trans−cis Switching Mechanisms in Proline Analogues and Their Relevance for the Gating of the 5-HT3 Receptor. The Journal of Physical Chemistry B 113: 12148-12153.

158. Paulsen IM, Martin IL, Dunn SMJ (2009) Isomerization of the proline in the M2-M3 linker is not required for activation of the human 5-HT3A receptor. Journal of Neurochemistry 110: 870-878.

159. Cheng X, Ivanov I, Wang H, Sine SM, McCammon JA (2007) Nanosecond-Timescale Conformational Dynamics of the Human α7 Nicotinic Acetylcholine Receptor. Biophysical Journal 93: 2622-2634.

160. Keramidas A, Lynch JW (2013) An outline of desensitization in pentameric ligand-gated ion channel receptors. Cellular and Molecular Life Sciences 70: 1241-1253.

161. Kash TL, Jenkins A, Kelley JC, Trudell JR, Harrison NL (2003) Coupling of agonist binding to channel gating in the GABAA receptor. Nature 421: 272-275.

162. Zhu F, Hummer G (2010) Pore opening and closing of a pentameric ligand-gated ion channel. Proceedings of the National Academy of Sciences of the United States of America 107: 19814-19819.

163. Galzi JL, Devillers-Thiéry A, Hussy N, Bertrand S, Changeux JP, Bertrand D (1992) Mutations in the channel domain of a neuronal nicotinic receptor convert ion selectivity from cationic to anionic. Nature 359: 500-505.

164. Corringer P-J, Bertrand S, Galzi J-L, Devillers-Thiéry A, Changeux J-P, Bertrand D (1999) Mutational analysis of the charge selectivity filter of the α7 nicotinic acetylcholine receptor. Neuron 22: 831-843.

165. Keramidas A, Moorhouse AJ, Pierce KD, Schofield PR, Barry PH (2002) Cation-selective mutations in the M2 domain of the inhibitory glycine receptor channel reveal determinants of ion-charge selectivity. The Journal of General Physiology 119: 393-410.

166. Wotring VE, Miller TS, Weiss DS (2003) Mutations at the GABA receptor selectivity filter: a possible role for effective charges. The Journal of Physiology 548: 527-540.

167. Peters JA, Cooper MA, Carland JE, Livesey MR, Hales TG, Lambert JJ (2010) Novel structural determinants of single channel conductance and ion selectivity in 5-hydroxytryptamine type 3 and nicotinic acetylcholine receptors. The Journal of Physiology 588: 587-596.

168. Hansen SB, Wang HL, Taylor P, Sine SM (2008) An Ion Selectivity Filter in the Extracellular Domain of Cys-loop Receptors Reveals Determinants for Ion Conductance. Journal of Biological Chemistry 283: 36066-36070.

169. Barnes NM, Hales TG, Lummis SC, Peters JA (2009) The 5-HT3 receptor--the relationship between structure and function. Neuropharmacology 56: 273-284.

170. Lummis SCR (2012) 5-HT3 Receptors. Journal of Biological Chemistry 287: 40239-40245.

171. Walstab J, Rappold G, Niesler B (2010) 5-HT3 receptors: Role in disease and target of drugs. Pharmacology & Therapeutics 128: 146-169.

172. Beene DL, Brandt GS, Zhong W, Zacharias NM, Lester HA, Dougherty DA (2002) Cation-pi interactions in ligand recognition by serotonergic (5-HT3A) and nicotinic acetylcholine receptors: the anomalous binding properties of nicotine. Biochemistry 41: 10262-10269.

173. Price KL, Bower KS, Thompson AJ, Lester HA, Dougherty DA, Lummis SCR (2008) A Hydrogen Bond in Loop A Is Critical for the Binding and Function of the 5-HT3 Receptor†. Biochemistry 47: 6370-6377.

174. Sullivan NL, Sullivan NL, Thompson AJ, Price KL, Lummis SCR (2006) Defining the roles of Asn-128, Glu-129 and Phe-130 in loop A of the 5-HT3 receptor. Molecular Membrane Biology 23: 442-451.

175. Beene DL, Price KL, Lester HA, Dougherty DA, Lummis SC (2004) Tyrosine residues that control binding and gating in the 5-hydroxytryptamine3 receptor revealed by unnatural amino acid mutagenesis. J Neurosci 24: 9097-9104.

176. Thompson AJ, Lochner M, Lummis SCR (2008) Loop B Is a Major Structural Component of the 5-HT3 Receptor. Biophysical Journal 95: 5728-5736.

177. Thompson AJ, Price KL, Reeves DC, Chan SL, Chau PL, Lummis SCR (2005) Locating an Antagonist in the 5-HT3 Receptor Binding Site Using Modeling and Radioligand Binding. Journal of Biological Chemistry 280: 20476-20482.

178. Reeves DC, Sayed MFR, Chau P-L, Price KL, Lummis SCR (2003) Prediction of 5-HT3 Receptor Agonist-Binding Residues Using Homology Modeling. Biophysical journal 84: 2338-2344.

179. Thompson AJ, Padgett CL, Lummis SCR (2006) Mutagenesis and Molecular Modeling Reveal the Importance of the 5-HT3 Receptor F-loop. Journal of Biological Chemistry 281: 16576-16582.

180. Thompson AJ, Chau PL, Chan SL, Lummis SCR (2006) Unbinding Pathways of an Agonist and an Antagonist from the 5-HT3 Receptor. Biophysical Journal 90: 1979-1991.

181. Panicker S, Cruz H, Arrabit C, Slesinger PA (2002) Evidence for a centrally located gate in the pore of a serotonin-gated ion channel. The Journal of neuroscience 22: 1629-1639.

182. Carland JE, Cooper MA, Livesey MR, Hales TG, Peters JA, Lambert JJ (2013) Mutagenic Analysis of the Intracellular Portals of the Human 5-HT3A Receptor. Journal of Biological Chemistry 288: 31592-31601.

183. Gunthorpe MJ (2001) Conversion of the Ion Selectivity of the 5-HT3A Receptor from Cationic to Anionic Reveals a Conserved Feature of the Ligand-gated Ion Channel Superfamily. Journal of Biological Chemistry 276: 10977-10983.

184. Thompson AJ, Lummis SCR (2003) A single ring of charged amino acids at one end of the pore can control ion selectivity in the 5-HT3 receptor. British Journal of Pharmacology 140: 359-365.

185. Kelley SP, Dunlop JI, Kirkness EF, Lambert JJ, Peters JA (2003) A cytoplasmic region determines single-channel conductance in 5-HT3 receptors. Nature 424: 321-324.

186. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Research 40: D370-376.

187. Jo S, Lim JB, Klauda JB, Im W (2009) CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. Biophysical Journal 97: 50-58.

188. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. Journal of Molecular Graphics 14: 33-38, 27-28.

189. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology 234: 779-815.

190. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. (2005) Scalable molecular dynamics with NAMD. Journal of Computational Chemistry 26: 1781-1802.

191. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. The Journal of Physical Chemistry B 102: 3586-3616.

192. Caplan DA, Subbotina JO, Noskov SY (2008) Molecular Mechanism of Ion-Ion and Ion-Substrate Coupling in the Na+-Dependent Leucine Transporter LeuT. Biophysical Journal 95: 4613-4621.

193. Noskov SY, Roux B (2008) Control of Ion Selectivity in LeuT: Two Na+ Binding Sites with Two Different Mechanisms. Journal of Molecular Biology 377: 804-818.

194. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, et al. (2010) Update of the CHARMM All-Atom Additive

Force Field for Lipids: Validation on Six Lipid Types. The Journal of Physical Chemistry B 114: 7830-7843.

195. Smart OS, Neduvelil JG, Wang X, Wallace BA, Sansom MS (1996) HOLE: a program for the analysis of the pore dimensions of ion channel structural models. Journal of Molecular Graphics 14: 354-360, 376.

196. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. Journal of Computational Chemistry 32: 2319-2327.

197. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. Journal of Computational Chemistry 25: 1605-1612.

198. Hunter JD (2007) Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering 9: 90-95.

199. Unwin N (2005) Refined Structure of the Nicotinic Acetylcholine Receptor at 4Å Resolution. Journal of Molecular Biology 346: 967-989.

200. Althoff T, Hibbs RE, Banerjee S, Gouaux E (2014) X-ray structures of GluCl in apo states reveal a gating mechanism of Cys-loop receptors. Nature 512: 333-337.

201. Pan J, Chen Q, Willenbring D, Yoshida K, Tillman T, Kashlan OB, et al. (2012) Structure of the pentameric ligand-gated ion channel ELIC cocrystallized with its competitive antagonist acetylcholine. Nature Communications 3.

202. Nury H, Van Renterghem C, Weng Y, Tran A, Baaden M, Dufresne V, et al. (2011) X-ray structures of general anaesthetics bound to a pentameric ligand-gated ion channel. Nature 469: 428-431.

203. Wang H-L, Cheng X, Taylor P, McCammon JA, Sine SM (2008) Control of Cation Permeation through the Nicotinic Receptor Channel. PLoS Computational Biology 4.

204. Wang H-L, Toghraee R, Papke D, Cheng X-L, McCammon JA, Ravaioli U, et al. (2009) Single-Channel Current Through Nicotinic Receptor Produced by Closure of Binding Site C-Loop. Biophysical Journal 96: 3582-3590.

205. Sauguet L, Poitevin F, Murail S, Van Renterghem C, Moraga-Cid G, Malherbe L, et al. (2013) Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels. The EMBO journal 32: 728-741.

206. Hales TG, Dunlop JI, Deeb TZ, Carland JE, Kelley SP, Lambert JJ, et al. (2006) Common Determinants of Single Channel Conductance within the Large Cytoplasmic Loop of 5-Hydroxytryptamine Type 3 and 4beta2 Nicotinic Acetylcholine Receptors. Journal of Biological Chemistry 281: 8062-8071.

207. Bouzat C, Bartos M, Corradi J, Sine SM (2008) The Interface between Extracellular and Transmembrane Domains of Homomeric Cys-Loop Receptors Governs Open-Channel Lifetime and Rate of Desensitization. Journal of Neuroscience 28: 7808-7819.

208. Reeves DC (2005) A Role for the β1- β2 Loop in the Gating of 5-HT3 Receptors. Journal of Neuroscience 25: 9358-9366.

209. Cheng X, Wang H, Grant B, Sine SM, McCammon JA (2006) Targeted Molecular Dynamics Study of C-Loop Closure and Channel Gating in Nicotinic Receptors. PLoS Computational Biology 2.

210. Gao F, Bren N, Burghardt TP, Hansen S, Henchman RH, Taylor P, et al. (2005) Agonist-mediated Conformational Changes in Acetylcholine-binding Protein Revealed by Simulation and Intrinsic Tryptophan Fluorescence. Journal of Biological Chemistry 280: 8443-8451.

211. Henchman RH, Wang H-L, Sine SM, Taylor P, Andrew McCammon J (2005) Ligand-Induced Conformational Change in the α7 Nicotinic Receptor Ligand Binding Domain. Biophysical Journal 88: 2564-2576.

212. Liu X, Xu Y, Li H, Wang X, Jiang H, Barrantes FJ (2008) Mechanics of channel gating of the nicotinic acetylcholine receptor. PLoS computational biology 4.

213. Yi M, Tjong H, Zhou H-X (2008) Spontaneous conformational change and toxin binding in α7 acetylcholine receptor: Insight into channel activation and inhibition. Proceedings of the National Academy of Sciences 105: 8280-8285.

214. Reeves DC (2001) Structural and Electrostatic Properties of the 5-HT3 Receptor Pore Revealed by Substituted Cysteine Accessibility Mutagenesis. Journal of Biological Chemistry 276: 42035-42042.

215. Livesey MR, Cooper MA, Lambert JJ, Peters JA (2011) Rings of Charge within the Extracellular Vestibule Influence Ion Permeation of the 5-HT3A Receptor. Journal of Biological Chemistry 286: 16008-16017.

216. Jansen M, Bali M, Akabas MH (2008) Modular Design of Cys-loop Ligand-gated Ion Channels: Functional 5-HT3 and GABA 1 Receptors Lacking the Large Cytoplasmic M3M4 Loop. The Journal of General Physiology 131: 137-146.

217. Livesey MR, Cooper MA, Deeb TZ, Carland JE, Kozuska J, Hales TG, et al. (2008) Structural Determinants of Ca2+ Permeability and Conduction in the Human 5-Hydroxytryptamine Type 3A Receptor. Journal of Biological Chemistry 283: 19301-19313.

218. Hibbs RE, Gouaux E (2011) Principles of activation and permeation in an anion-selective Cys-loop receptor. Nature 474: 54-60.

219. Bernèche S, Roux B (2001) Energetics of ion conduction through the K+ channel. Nature 414: 73-77.

220. Hilf RJC, Dutzler R (2008) X-ray structure of a prokaryotic pentameric ligand-gated ion channel. Nature 452: 375-379.

221. Deeb TZ, Carland JE, Cooper MA, Livesey MR, Lambert JJ, Peters JA, et al. (2006) Dynamic Modification of a Mutant Cytoplasmic Cysteine Residue Modulates the Conductance of the Human 5-HT3A Receptor. Journal of Biological Chemistry 282: 6172-6182.

222. Kozuska JL, Paulsen IM, Belfield WJ, Martin IL, Cole DJ, Holt A, et al. (2014) Impact of intracellular domain flexibility upon properties of activated human 5-HT3 receptors: 5-HT3 receptor intracellular domain significance. British Journal of Pharmacology 171: 1617-1628.

223. Ilegems E, Pick H, Deluz C, Kellenberger S, Vogel H (2005) Ligand Binding Transmits Conformational Changes across the Membrane-Spanning Region to the Intracellular Side of the 5-HT3 Serotonin Receptor. ChemBioChem 6: 2180-2185.

224. Blount P, Moe PC (1999) Bacterial mechanosensitive channels: integrating physiology, structure and function. Trends in microbiology 7: 420-424.

225. Booth IR, Blount P (2012) The MscS and MscL Families of Mechanosensitive Channels Act as Microbial Emergency Release Valves. Journal of Bacteriology 194: 4802-4809.

226. Steinbacher S, Bass R, Strop P, Rees DC (2007) Structures of the Prokaryotic Mechanosensitive Channels MscL and MscS. In: Hamill OP, editor. Current Topics in Membranes: Academic Press. pp. 1-24.

227. Ou X, Blount P, Hoffman RJ, Kung C (1998) One face of a transmembrane helix is crucial in mechanosensitive channel gating. Proceedings of the National Academy of Sciences of the United States of America 95: 11471-11475.

228. Anishkin A, Chiang C-S, Sukharev S (2005) Gain-of-function mutations reveal expanded intermediate states and a sequential action of two gates in MscL. The Journal of general physiology 125: 155-170.

229. Yoshimura K, Batiza A, Schroeder M, Blount P, Kung C (1999) Hydrophilicity of a Single Residue within MscL Correlates with Increased Channel Mechanosensitivity. Biophysical Journal 77: 1960-1972.

230. Bartlett JL, Li Y, Blount P (2006) Mechanosensitive Channel Gating Transitions Resolved by Functional Changes upon Pore Modification. Biophysical Journal 91: 3684-3691.

231. Batiza AF, Kuo MM-C, Yoshimura K, Kung C (2002) Gating the bacterial mechanosensitive channel MscL in vivo. Proceedings of the National Academy of Sciences 99: 5643-5648.

232. Yoshimura K, Batiza A, Kung C (2001) Chemically Charging the Pore Constriction Opens the Mechanosensitive Channel MscL. Biophysical Journal 80: 2198-2206.

233. Doerner JF, Febvay S, Clapham DE (2012) Controlled delivery of bioactive molecules into live cells using the bacterial mechanosensitive channel MscL. Nature Communications 3.

234. Koçer A, Walko M, Bulten E, Halza E, Feringa BL, Meijberg W (2006) Rationally Designed Chemical Modulators Convert a Bacterial Channel

Protein into a pH-Sensory Valve. Angewandte Chemie International Edition 45: 3126-3130.

235. Birkner JP, Poolman B, Koçer A (2012) Hydrophobic gating of mechanosensitive channel of large conductance evidenced by single-subunit resolution. Proceedings of the National Academy of Sciences of the United States of America 109: 12944-12949.

236. Mika JT, Birkner JP, Poolman B, Koçer A (2013) On the role of individual subunits in MscL gating: "all for one, one for all?". FASEB journal : official publication of the Federation of American Societies for Experimental Biology 27: 882-892.

237. Colombo G, Marrink SJ, Mark AE (2003) Simulation of MscL gating in a bilayer under stress. Biophysical journal 84: 2331-2337.

238. Gullingsrud J, Schulten K (2004) Lipid Bilayer Pressure Profiles and Mechanosensitive Channel Gating. Biophysical Journal 86: 3496-3509.

239. Jeon J, Voth GA (2008) Gating of the Mechanosensitive Channel Protein MscL: The Interplay of Membrane and Protein. Biophysical Journal 94: 3497-3511.

240. Tang Y, Yoo J, Yethiraj A, Cui Q, Chen X (2008) Gating Mechanisms of Mechanosensitive Channels of Large Conductance, II: Systematic Study of Conformational Transitions. Biophysical Journal 95: 581-596.

241. Yefimov S, van der Giessen E, Onck PR, Marrink SJ (2008) Mechanosensitive membrane channels in action. Biophysical journal 94: 2994-3002.

242. Elmore DE, Dougherty DA (2003) Investigating lipid composition effects on the mechanosensitive channel of large conductance (MscL) using molecular dynamics simulations. Biophysical journal 85: 1512-1524.

243. Meyer GR, Gullingsrud J, Schulten K, Martinac B (2006) Molecular dynamics study of MscL interactions with a curved lipid bilayer. Biophysical journal 91: 1630-1637.

244. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, et al. (2010) Update of the CHARMM all-atom additive force

field for lipids: validation on six lipid types. The journal of physical chemistry B 114: 7830-7843.

245. MacKerell AD, Bashford D, Bellott, Dunbrack RL, Evanseck JD, Field MJ, et al. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. The Journal of Physical Chemistry B 102: 3586-3616.

246. Mackerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. Journal of Computational Chemistry 25: 1400-1415.

247. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. (2009) Gaussian 09, Revision B.01.

248. Dupradeau F-Y, Pigache A, Zaffran T, Savineau C, Lelong R, Grivel N, et al. (2010) The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. Physical Chemistry Chemical Physics 12: 7821-7839.

249. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of Computational Chemistry 31: 671-690.

250. Pezeshki S, Chimerel C, Bessonov AN, Winterhalter M, Kleinekathöfer U (2009) Understanding ion conductance on a molecular level: an all-atom modeling of the bacterial porin OmpF. Biophysical journal 97: 1898-1906.

251. Aksimentiev A, Schulten K (2005) Imaging alpha-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. Biophysical journal 88: 3745-3761.

252. Wang Y, Liu Y, DeBerg HA, Nomura T, Hoffman MT, Rohde PR, et al. (2014) Single molecule FRET reveals pore size and opening mechanism of a mechano-sensitive ion channel. eLife 3.

253. Cruickshank CC, Minchin RF, Le Dain AC, Martinac B (1997) Estimation of the pore size of the large-conductance mechanosensitive ion channel of Escherichia coli. Biophysical Journal 73: 1925-1931.

254. Sotomayor M, Schulten K (2004) Molecular dynamics study of gating in the mechanosensitive channel of small conductance MscS. Biophysical journal 87: 3050-3065.

255. Gullingsrud J, Kosztin D, Schulten K (2001) Structural determinants of MscL gating studied by molecular dynamics simulations. Biophysical journal 80: 2074-2081.

256. Beckstein O, Biggin PC, Sansom MSP (2001) A Hydrophobic Gating Mechanism for Nanopores. The Journal of Physical Chemistry B 105: 12902-12905.

257. Beckstein O, Sansom MSP (2004) The influence of geometry, surface character, and flexibility on the permeation of ions and water through biological pores. Physical biology 1: 42-52.

258. Oakley AJ, Martinac B, Wilce MC (1999) Structure and function of the bacterial mechanosensitive channel of large conductance. Protein science : a publication of the Protein Society 8: 1915-1921.

259. Martinac B (2011) Bacterial Mechanosensitive Channels as a Paradigm for Mechanosensory Transduction. Cellular Physiology and Biochemistry 28: 1051-1060.

260. Maurer JA, Elmore DE, Lester HA, Dougherty DA (2000) Comparing and contrasting Escherichia coli and Mycobacterium tuberculosis mechanosensitive channels (MscL). New gain of function mutations in the loop region. The Journal of biological chemistry 275: 22238-22244.

261. Moe PC, Levin G, Blount P (2000) Correlating a protein structure with function of a bacterial mechanosensitive channel. The Journal of biological chemistry 275: 31121-31127.

262. Corry B, Hurst AC, Pal P, Nomura T, Rigby P, Martinac B (2010) An improved open-channel structure of MscL determined from FRET confocal microscopy and simulation. The Journal of general physiology 136: 483-494.

263. Perozo E, Cortes DM, Sompornpisut P, Kloda A, Martinac B (2002) Open channel structure of MscL and the gating mechanism of mechanosensitive channels. Nature 418: 942-948.

264. Perozo E, Kloda A, Cortes DM, Martinac B (2002) Physical principles underlying the transduction of bilayer deformation forces during mechanosensitive channel gating. Nature structural biology 9: 696-703.

265. Sukharev S, Durell SR, Guy HR (2001) Structural models of the MscL gating mechanism. Biophysical journal 81: 917-936.

266. Louhivuori M, Risselada HJ, van der Giessen E, Marrink SJ (2010) Release of content through mechano-sensitive gates in pressurized liposomes. Proceedings of the National Academy of Sciences of the United States of America 107: 19856-19860.

267. Iscla I, Levin G, Wray R, Blount P (2007) Disulfide trapping the mechanosensitive channel MscL into a gating-transition state. Biophysical journal 92: 1224-1232.

268. Shirvanyants D, Ramachandran S, Mei Y, Xu L, Meissner G, Dokholyan NV (2014) Pore Dynamics and Conductance of RyR1 Transmembrane Domain. Biophysical Journal 106: 2375-2384.

269. Zhong D, Blount P (2013) Phosphatidylinositol Is Crucial for the Mechanosensitivity of Mycobacterium tuberculosis MscL. Biochemistry 52: 5415-5420.

270. Deplazes E, Louhivuori M, Jayatilaka D, Marrink SJ, Corry B (2012) Structural investigation of MscL gating using experimental data and coarse grained MD simulations. PLoS computational biology 8: e1002683-e1002683.

271. Gullingsrud J, Schulten K (2003) Gating of MscL studied by steered molecular dynamics. Biophysical journal 85: 2087-2099.

272. Kong Y, Shen Y, Warth TE, Ma J (2002) Conformational pathways in the gating of Escherichia coli mechanosensitive channel. Proceedings of the National Academy of Sciences of the United States of America 99: 5999-6004.

273. Nestorovich EM, Bezrukov SM (2012) Obstructing Toxin Pathways by Targeted Pore Blockage. Chemical Reviews 112: 6388-6430.

274. Gurnev PA, Nestorovich EM (2014) Channel-forming bacterial toxins in biosensing and macromolecule delivery. Toxins 6: 2483-2540.

275. Menestrina G, Serra MD, Pre G (2001) Mode of action of b -barrel pore-forming toxins of the staphylococcal a -hemolysin family. 39.

276. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. (2008) The potential and challenges of nanopore sequencing. Nature Biotechnology 26: 1146-1153.

277. Jetha NNW, W.; Marziali A. (2009) Forming an α-Hemolysin Nanopore for Single-Molecule Analysis - Springer. In: Foote RS, Lee JW, editors: Humana Press.

278. Bayley H, Cremer PS (2001) Stochastic sensors inspired by biology. Nature 413: 226-230.

279. Merino E, Ribagorda M (2012) Control over molecular motion using the cis-trans photoisomerization of the azo group. Beilstein journal of organic chemistry 8: 1071-1090.

280. Banghart MR, Volgraf M, Trauner D, Matthew R. Banghart MVDT (2006) Engineering Light-Gated Ion Channels. pp. 15129-15141.

281. Jog PV, Gin MS (2008) LETTERS A Light-Gated Synthetic Ion Channel. 9-12.

282. McCullagh M, Franco I, Ratner MA, Schatz GC (2011) DNA-based optomechanical molecular motor. Journal of the American Chemical Society 133: 3452-3459.

283. Böckmann M, Peter C, Site LD, Doltsinis NL, Kremer K, Marx D (2007) Atomistic Force Field for Azobenzene Compounds Adapted for QM/MM Simulations with Applications to Liquids and Liquid Crystals. Journal of Chemical Theory and Computation 3: 1789-1802.

284. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics 23: 327-341.

285. Martin HSC, Jha S, Coveney PV (2014) Comparative analysis of nucleotide translocation through protein nanopores using steered molecular dynamics and an adaptive biasing force. Journal of computational chemistry 35: 692-702.

286. Banghart M, Borges K, Isacoff E, Trauner D, Kramer RH (2004) Light-activated ion channels for remote control of neuronal firing. Nature Neuroscience 7: 1381-1386.

287. Chalker JM, Bernardes GaJL, Davis BG (2011) A "Tag-and-Modify" Approach to Site-Selective Protein Modification. Accounts of Chemical Research 44: 730-741.

288. Ying Y-L, Zhang J, Meng F-N, Cao C, Yao X, Willner I, et al. (2013) A stimuli-responsive nanopore based on a photoresponsive host-guest system. Scientific reports 3: 1662-1662.

289. Py B, Barras F (2010) Building Fe–S proteins: bacterial strategies. Nature Reviews Microbiology 8: 436-446.

290. Isaya G (2014) Mitochondrial iron-sulfur cluster dysfunction in neurodegenerative disease. Frontiers in Pharmacology 5.

291. Hadzhieva M, Kirches E, Mawrin C (2014) Review: Iron metabolism and the role of iron in neurodegenerative disorders: Iron in neurodegeneration. Neuropathology and Applied Neurobiology 40: 240-257.

292. Campuzano V, Montermini L, Moltò MD, Pianese L, Cossée M, Cavalcanti F, et al. (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science (New York, NY) 271: 1423-1427.

293. Kollberg G, Tulinius M, Melberg A, Darin N, Andersen O, Holmgren D, et al. (2009) Clinical manifestation and a new ISCU mutation in iron-sulphur cluster deficiency myopathy. Brain: A Journal of Neurology 132: 2170-2179.

294. Mochel F, Knight MA, Tong W-H, Hernandez D, Ayyad K, Taivassalo T, et al. (2008) Splice mutation in the iron-sulfur cluster scaffold protein ISCU causes myopathy with exercise intolerance. American Journal of Human Genetics 82: 652-660.

295. Olsson A, Lind L, Thornell L-E, Holmberg M (2008) Myopathy with lactic acidosis is linked to chromosome 12q23.3-24.11 and caused by an intron mutation in the ISCU gene resulting in a splicing defect. Human Molecular Genetics 17: 1666-1672.

296. Calvo SE, Tucker EJ, Compton AG, Kirby DM, Crawford G, Burtt NP, et al. (2010) High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. Nature Genetics 42: 851-858.

297. Camaschella C, Campanella A, De Falco L, Boschetto L, Merlini R, Silvestri L, et al. (2007) The human counterpart of zebrafish shiraz shows sideroblastic-like microcytic anemia and iron overload. Blood 110: 1353-1358.

298. Cameron JM, Janer A, Levandovskiy V, Mackay N, Rouault TA, Tong W-H, et al. (2011) Mutations in iron-sulfur cluster scaffold genes NFU1 and BOLA3 cause a fatal deficiency of multiple respiratory chain and 2-oxoacid dehydrogenase enzymes. American Journal of Human Genetics 89: 486-495.

299. Navarro-Sastre A, Tort F, Stehling O, Uzarska MA, Arranz JA, Del Toro M, et al. (2011) A fatal mitochondrial disease is associated with defective NFU1 function in the maturation of a subset of mitochondrial Fe-S proteins. American Journal of Human Genetics 89: 656-667.

300. Haack TB, Rolinski B, Haberberger B, Zimmermann F, Schum J, Strecker V, et al. (2013) Homozygous missense mutation in BOLA3 causes multiple mitochondrial dysfunctions syndrome in two siblings. Journal of Inherited Metabolic Disease 36: 55-62.

301. Tokumoto U, Takahashi Y (2001) Genetic analysis of the isc operon in Escherichia coli involved in the biogenesis of cellular iron-sulfur proteins. Journal of Biochemistry 130: 63-71.

302. Ramelot TA, Cort JR, Goldsmith-Fischman S, Kornhaber GJ, Xiao R, Shastry R, et al. (2004) Solution NMR Structure of the Iron–Sulfur Cluster Assembly Protein U (IscU) with Zinc Bound at the Active Site. Journal of Molecular Biology 344: 567-583.

303. Liu J, Oganesyan N, Shin D-H, Jancarik J, Yokota H, Kim R, et al. (2005) Structural characterization of an iron-sulfur cluster assembly protein IscU in a zinc-bound form. Proteins: Structure, Function, and Bioinformatics 59: 875-881.

304. Kim JH, Tonelli M, Kim T, Markley JL (2012) Three-Dimensional Structure and Determinants of Stability of the Iron–Sulfur Cluster Scaffold Protein IscU from Escherichia coli. Biochemistry 51: 5557-5563.

305. Shimomura Y, Wada K, Fukuyama K, Takahashi Y (2008) The Asymmetric Trimeric Architecture of [2Fe–2S] IscU: Implications for Its Scaffolding during Iron–Sulfur Cluster Biosynthesis. Journal of Molecular Biology 383: 133-143.

306. Cupp-Vickery JR, Urbina H, Vickery LE (2003) Crystal Structure of IscS, a Cysteine Desulfurase from Escherichia coli. Journal of Molecular Biology 330: 1049-1059.

307. Shi R, Proteau A, Villarroya M, Moukadiri I, Zhang L, Trempe J-F, et al. (2010) Structural Basis for Fe–S Cluster Assembly and tRNA Thiolation Mediated by IscS Protein–Protein Interactions. PLoS Biology 8: e1000354.

308. Marinoni EN, de Oliveira JS, Nicolet Y, Raulfs EC, Amara P, Dean DR, et al. (2012) (IscS-IscU)2 Complex Structures Provide Insights into Fe2S2 Biogenesis and Transfer. Angewandte Chemie International Edition 51: 5439-5442.

309. Yamanaka Y, Zeppieri L, Nicolet Y, Marinoni EN, de Oliveira JS, Odaka M, et al. (2013) Crystal structure and functional studies of an unusual cysteine desulfurase from Archaeoglobus fulgidus. Dalton Trans 42: 3092-3099.

310. Cho SJ, Lee MG, Yang JK, Lee JY, Song HK, Suh SW (2000) Crystal structure of Escherichia coli CyaY protein reveals a previously unidentified fold for the evolutionarily conserved frataxin family. Proceedings of the National Academy of Sciences of the United States of America 97: 8932-8937.

311. Nair M, Adinolfi S, Pastore C, Kelly G, Temussi P, Pastore A (2004) Solution structure of the bacterial frataxin ortholog, CyaY: mapping the iron binding sites. Structure (London, England: 1993) 12: 2037-2048.

312. Karlberg T, Schagerlöf U, Gakh O, Park S, Ryde U, Lindahl M, et al. (2006) The structures of frataxin oligomers reveal the mechanism for the delivery and detoxification of iron. Structure (London, England: 1993) 14: 1535-1546.

313. Bridwell-Rabb J, Winn AM, Barondeau DP (2011) Structure-function analysis of Friedreich's ataxia mutants reveals determinants of frataxin binding and activation of the Fe-S assembly complex. Biochemistry 50: 7265-7274.

314. Söderberg CAG, Rajan S, Shkumatov AV, Gakh O, Schaefer S, Ahlgren E-C, et al. (2013) The molecular basis of iron-induced oligomerization of frataxin and the role of the ferroxidation reaction in oligomerization. The Journal of Biological Chemistry 288: 8156-8167.

315. Pastore C, Adinolfi S, Huynen MA, Rybin V, Martin S, Mayer M, et al. (2006) YfhJ, a molecular adaptor in iron-sulfur cluster formation or a frataxin-like protein? Structure (London, England: 1993) 14: 857-867.

316. Cupp-Vickery JR, Peterson JC, Ta DT, Vickery LE (2004) Crystal Structure of the Molecular Chaperone HscA Substrate Binding Domain Complexed with the IscU Recognition Peptide ELPPVKIHC. Journal of Molecular Biology 342: 1265-1278.

317. Kakuta Y, Horio T, Takahashi Y, Fukuyama K (2001) Crystal Structure of Escherichia coli Fdx, an Adrenodoxin-Type Ferredoxin Involved in the Assembly of Iron−Sulfur Clusters. Biochemistry 40: 11007-11012.

318. Cupp-Vickery JR, Silberg JJ, Ta DT, Vickery LE (2004) Crystal Structure of IscA, an Iron-sulfur Cluster Assembly Protein from Escherichia coli. Journal of Molecular Biology 338: 127-137.

319. Füzéry AK, Tonelli M, Ta DT, Cornilescu G, Vickery LE, Markley JL (2008) Solution Structure of the Iron−Sulfur Cluster Cochaperone HscB and Its Binding Surface for the Iron−Sulfur Assembly Scaffold Protein IscU. Biochemistry 47: 9394-9404.

320. Maio N, Rouault TA (2015) Iron–sulfur cluster biogenesis in mammalian cells: New insights into the molecular mechanisms of cluster delivery. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1853: 1493-1512.

321. Biederbick A, Stehling O, Rösser R, Niggemeyer B, Nakai Y, Elsässer H-P, et al. (2006) Role of human mitochondrial Nfs1 in cytosolic iron-sulfur protein biogenesis and iron regulation. Molecular and Cellular Biology 26: 5675-5687.

322. Léon S, Touraine B, Briat J-F, Lobréaux S (2005) Mitochondrial localization of Arabidopsis thaliana Isu Fe-S scaffold proteins. FEBS Letters 579: 1930-1934.

323. Schwartz CJ, Djaman O, Imlay JA, Kiley PJ (2000) The cysteine desulfurase, IscS, has a major role in in vivo Fe-S cluster formation in Escherichia coli. Proceedings of the National Academy of Sciences 97: 9009-9014.

324. Marelja Z, Stöcklein W, Nimtz M, Leimkühler S (2008) A novel role for human Nfs1 in the cytoplasm: Nfs1 acts as a sulfur donor for MOCS3, a protein involved in molybdenum cofactor biosynthesis. The Journal of Biological Chemistry 283: 25178-25185.

325. Paris Z, Changmai P, Rubio MAT, Zikova A, Stuart KD, Alfonzo JD, et al. (2010) The Fe/S Cluster Assembly Protein Isd11 Is Essential for tRNA Thiolation in Trypanosoma brucei. Journal of Biological Chemistry 285: 22394-22402.

326. Lill R (2009) Function and biogenesis of iron–sulphur proteins. Nature 460: 831-838.

327. Agar JN, Krebs C, Frazzon J, Huynh BH, Dean DR, Johnson MK (2000) IscU as a Scaffold for Iron−Sulfur Cluster Biosynthesis: Sequential Assembly of [2Fe-2S] and [4Fe-4S] Clusters in IscU. Biochemistry 39: 7856-7862.

328. Bonomi F, Iametti S, Morleo A, Ta D, Vickery LE (2011) Facilitated Transfer of IscU–[2Fe2S] Clusters by Chaperone-Mediated Ligand Exchange. Biochemistry 50: 9641-9650.

329. Markley JL, Kim JH, Dai Z, Bothe JR, Cai K, Frederick RO, et al. (2013) Metamorphic protein IscU alternates conformations in the course of its role as the scaffold protein for iron-sulfur cluster biosynthesis and delivery. FEBS Letters 587: 1172-1179.

330. Iannuzzi C, Adrover M, Puglisi R, Yan R, Temussi PA, Pastore A (2014) The role of zinc in the stability of the marginally stable IscU scaffold protein. Protein Science: A Publication of the Protein Society 23: 1208-1219.

331. Roche B, Aussel L, Ezraty B, Mandin P, Py B, Barras F (2013) Iron/sulfur proteins biogenesis in prokaryotes: Formation, regulation and diversity. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1827: 455-469.

332. Adinolfi S, Iannuzzi C, Prischi F, Pastore C, Iametti S, Martin SR, et al. (2009) Bacterial frataxin CyaY is the gatekeeper of iron-sulfur cluster formation catalyzed by IscS. Nature Structural & Molecular Biology 16: 390-396.

333. Prischi F, Konarev PV, Iannuzzi C, Pastore C, Adinolfi S, Martin SR, et al. (2010) Structural bases for the interaction of frataxin with the central components of iron–sulphur cluster assembly. Nature Communications 1: 95.

334. Bridwell-Rabb J, Iannuzzi C, Pastore A, Barondeau DP (2012) Effector Role Reversal during Evolution: The Case of Frataxin in Fe–S Cluster Biosynthesis. Biochemistry 51: 2506-2514.

335. Iannuzzi C, Adinolfi S, Howes BD, Garcia-Serres R, Clémancey M, Latour J-M, et al. (2011) The Role of CyaY in Iron Sulfur Cluster Assembly on the E. coli IscU Scaffold Protein. PLoS ONE 6: e21992.

336. Bridwell-Rabb J, Fox NG, Tsai C-L, Winn AM, Barondeau DP (2014) Human Frataxin Activates Fe–S Cluster Biosynthesis by Facilitating Sulfur Transfer Chemistry. Biochemistry 53: 4904-4913.

337. Colin F, Martelli A, Clémancey M, Latour J-M, Gambarelli S, Zeppieri L, et al. (2013) Mammalian Frataxin Controls Sulfur Production and Iron Entry during de Novo Fe4S4 Cluster Assembly. Journal of the American Chemical Society 135: 733-740.

338. Fox NG, Das D, Chakrabarti M, Lindahl PA, Barondeau DP (2015) Frataxin Accelerates [2Fe-2S] Cluster Formation on the Human Fe–S Assembly Complex. Biochemistry 54: 3880-3889.

339. Pandey A, Gordon DM, Pain J, Stemmler TL, Dancis A, Pain D (2013) Frataxin Directly Stimulates Mitochondrial Cysteine Desulfurase by Exposing Substrate-binding Sites, and a Mutant Fe-S Cluster Scaffold Protein with Frataxin-bypassing Ability Acts Similarly. Journal of Biological Chemistry 288: 36773-36786.

340. Parent A, Elduque X, Cornu D, Belot L, Le Caer J-P, Grandas A, et al. (2015) Mammalian frataxin directly enhances sulfur transfer of NFS1 persulfide to both ISCU and free thiols. Nature Communications 6: 5686.

341. Tsai C-L, Barondeau DP (2010) Human Frataxin Is an Allosteric Switch That Activates the Fe−S Cluster Biosynthetic Complex. Biochemistry 49: 9132-9139.

342. Andersen HC (1983) Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. Journal of Computational Physics 52: 24-34.

343. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. Journal of the American Chemical Society 117: 5179-5197.

344. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78: 1950-1958.

345. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. The Journal of Physical Chemistry 97: 10269-10280.

346. Vanquelef E, Simon S, Marquant G, Garcia E, Klimerak G, Delepine JC, et al. (2011) R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. Nucleic Acids Research 39: W511-517.

347. Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. Journal of Molecular Graphics & Modelling 25: 247-260.

348. Lindahl E, Hess B, van der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. Journal of Molecular Modeling 7: 306-317.

349. Yan R, Konarev PV, Iannuzzi C, Adinolfi S, Roche B, Kelly G, et al. (2013) Ferredoxin Competes with Bacterial Frataxin in Binding to the Desulfurase IscS. Journal of Biological Chemistry 288: 24777-24787.

350. Kim JH, Bothe JR, Frederick RO, Holder JC, Markley JL (2014) Role of IscX in iron-sulfur cluster biogenesis in Escherichia coli. Journal of the American Chemical Society 136: 7933-7942.

351. Schmucker S, Martelli A, Colin F, Page A, Wattenhofer-Donzé M, Reutenauer L, et al. (2011) Mammalian Frataxin: An Essential Function for Cellular Viability through an Interaction with a Preformed ISCU/NFS1/ISD11 Iron-Sulfur Assembly Complex. PLoS ONE 6: e16199.

352. Wang T, Craig EA (2008) Binding of Yeast Frataxin to the Scaffold for Fe-S Cluster Biogenesis, Isu. Journal of Biological Chemistry 283: 12674-12679.

353. Lauhon CT (2000) The iscS Gene in Escherichia coli Is Required for the Biosynthesis of 4-Thiouridine, Thiamin, and NAD. Journal of Biological Chemistry 275: 20096-20103.

354. Lauhon CT (2002) Requirement for IscS in Biosynthesis of All Thionucleosides in Escherichia coli. Journal of Bacteriology 184: 6820-6829.

355. Bedekovics T, Gajdos GB, Kispal G, Isaya G (2007) Partial conservation of functions between eukaryotic frataxin and the Escherichia coli frataxin homolog CyaY. FEMS yeast research 7: 1276-1284.

356. Cavadini P, Gellera C, Patel PI, Isaya G (2000) Human frataxin maintains mitochondrial iron homeostasis in Saccharomyces cerevisiae. Human Molecular Genetics 9: 2523-2530.

357. Adam AC, Bornhövd C, Prokisch H, Neupert W, Hell K (2006) The Nfs1 interacting protein Isd11 has an essential role in Fe/S cluster biogenesis in mitochondria. The EMBO journal 25: 174-183.

358. Teralı K, Beavil RL, Pickersgill RW, van der Giezen M (2013) The effect of the adaptor protein Isd11 on the quaternary structure of the eukaryotic cysteine desulphurase Nfs1. Biochemical and Biophysical Research Communications 440: 235-240.

359. Wiedemann N, Urzica E, Guiard B, Müller H, Lohaus C, Meyer HE, et al. (2006) Essential role of Isd11 in mitochondrial iron–sulfur cluster synthesis on Isu scaffold proteins. The EMBO Journal 25: 184-195.